Laboratori d’Enginyeria Mar´ıtima (LIM/UPC)
Universitat Polit`ecnica de Catalunya
Barcelona, Spain

[Twin Storm Events]

A probability analysis and risk evaluation of twin storm occurrences along the coast of Catalonia

Anna Maria Wojtanowicz

[07 June 2010]

Advisers:

Agustin Sanchez- Arcilla

Raimon Tolosana- Delgado

The concept of a *twin storm* refers to two sequential storm wave heights that are separated by a short period of calm water. In the past, these have often been regards as one storm with a ‘double peak’ or as two independent events.   This thesis will focus on determining a pattern of occurrence by analyzing the characteristics of the first storm and observing if these characteristics may have an influence on the occurrence of the second storm. Data has been taken from four areas along the Catalan coast and used to calculate probabilities of twin storm events based on maximum storm significant wave height, storm duration, and time of year of occurrence. Using these calculated probabilities, a risk analysis has been performed in order to decide if preventive action should be taken during the period of calm water in hopes of reducing the damage caused by the second storm. This thesis presents the results as well as conclusions that can be drawn from the analysis.

Table of Contents

List of Figures

List of Tables

Introduction

1.1 Thesis Objective

The coastline of Catalonia is experiencing an undeniable increase in population density as well as social and economical activity. Alongside this upward trend is the growing pressure on the coast to fulfill all the increasing needs required of it. In addition, the concerns associated to climate change are demanding that sustainability of the coast be guaranteed in the upcoming years in order for the region of Catalonia not to suffer losses due to unpreparedness. As well as decreasing the vulnerability of the coastal zone, fully understanding the potential hazards being faced are essential in the diminishment of risk. Once a clear understanding has been reached, future prediction becomes a goal which can hopefully lead to minimization of risk.

One of the hazards threatening the coastal zone is sea storms. An important characteristic of a sea storm is an increase in wave height. This increase, even slight, has the potential to be detrimental due to the quadratic relation between wave energy and wave height. Sea storms cause damage to man-built coastal structures as well as deform the natural composition of a coastal setting such as a beach or river outlet. Today, because of the high dependency on stable coastal features to allow for continuous port traffic for commerce or sufficient beach widths for tourism, understanding the pattern of wave storms has become increasingly important. This understanding of a pattern is the scope of this thesis, specifically focusing on the establishment of a *twin storm* pattern, which will be discussed in further detail later.

Once a pattern is recognized, the issue becomes to lessen the damage by taking preemptive measures. Determining when to do this depends on many factors such as the cost of these measures and the size of the damage. Setting up such factors into a risk analysis can aid in determining the best course of action.

The first objective of this thesis is to present a synopsis of the wave storm situation along the coast of Catalonia by taking buoy readings from the past eighteen years and categorizing them as storms based on maximum wave height, storm duration, and time of year in which they occur. The next point is to determine whether each storm has a twin. A storm containing a twin refers to a storm event followed by several hours of calm water which is then followed by another storm. Once all the storms are characterized with these values, the storms are implemented into a probability model so as to determine whether or not a pattern of occurrence exists. The next objective is to determine which variables influence whether or not a storm event has a twin by applying tests of independence. Determination of these variables is followed by a comparison of how each affect the outcome of a storm using the odds ratio and relative risks. This information is then applied to a risk analysis based on the performance of precautionary action or the absence of such action. Implementing this risk analysis can hopefully help to achieve the ultimate objective, namely, deciding when preventive action should or should not be taken.

Before undertaking the analysis, it must be noted that the concept of a twin storm is still a relatively new idea. Normal wave storms in the Mediterranean Sea are described as duration-limited because, on average, the duration last for less than 24 hours (Sanchez-Arcilla, Gonzalez-Marco, Bolanos, 2008). However, scientists have begun to observe a frequent characteristic of longer lasting storms, namely, the occurrence of *double peaks* (Jiménez, Sánchez, Valdemoro, Gracia, Nieto, 1997). A double peak event is a single storm with two large wave heights that are separated by calmer water. As implied by the definition, both wave heights were classified as the same storm event. The concept of a twin storm divides these two wave heights into two different events. This is done in order to explore the features of the first storm to determine a prediction method for the second storm. This type of categorization has not been researched before. The following analysis will shed some light on the topic, specifically focusing on recognizing a pattern and the factors that can account for the occurrence. As with all new research, there will be holes in the analysis and many of the results will be inconclusive due to lack of data. Therefore, this thesis also has the added objective to stimulate interest in the subject in order to motivate a more in-depth investigation.

1.2 Characteristics of the Catalan Coast

1.2.1 Geographical Information

The coast of Catalonia extends more than 580 kilometers from the France-Spain border to the Ebro Delta (Bowman, Guillén, López, Pellegrino, 2009). This stretch is characterized by both geological and geomorphic diversity throughout the length of the coast, specifically, features such as pocket beaches, headlands, cliffs, as well as long straight stretches of sandy beach containing sands with diameters that range from *0.2* to *0.6 mm*. It is considered to be a micro tidal, fetch-limited environment (Jimenez, 2009). One characteristic of the Catalan beaches is the absence of any dune ridge in the backshore of the beach. Erosion is a relevant problem along the coast with beach widths diminishing during storms. The following section will outline the beaches that make up the coast of Catalonia.

Based on coastal morphology, Mendoza (2008) classified the coast of Catalonia into seven areas. The northern-most area is *Costa Brava*, which is distinguished by its rocky coast, many cliffs, and short bay-beaches of course sand. To the south is the *Maresme*, a long stretch of course sand beginning at the Tordera River and stretching to Mongat. Both beaches are classified as reflective due to the course sand ($d_{50} > 0.6mm$) and steep slope of $\tan\beta \sim 0.1$(Mendoza and Jimenez, 2008). Southwest of Maresme is the *Barcelona* beach which is an artificial embayed beach of medium size sand extending from the Besos River to the Llobregat River. The *South Barcelona* beach is a nineteen km stretch of fine sandy beach starting in the Llobregat Delta and ending at the Port Ginesta. It is followed by the *Costa del Garraf*, a region containing low cliffs and pocket beaches which are mostly composed of fine sand. Next is the *Costa Dorada*, a fine sediment, low-lying coast that has a nearly-straight coastline. Finally, the Catalan coast ends at the *Ebro Delta*, where the Ebro River deposits fine sediments onto the mildly-slope, low-lying coast. These characteristics classify both the Costa Dorada and the Ebro Delta as dissipative beaches with sediment of size $0.2 < d_{50} < 0.25mm$ and beach slopes of $0.01 < \tan\beta < 0.02$ (Mendoza, Jimenez, 2008). The following figure illustrates the location of the seven coastal regions:

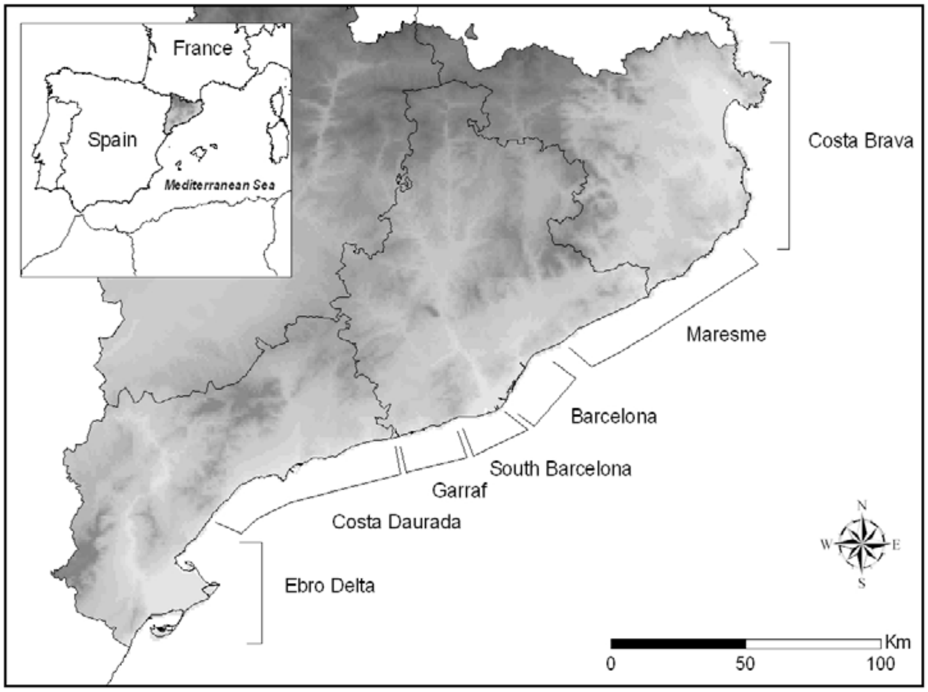

Figure 1: Classification of the Catalan Coast (Mendoza, 2008)

The coast of Catalonia is segregated by three of the four Catalan provinces: *Girona*, which contains the Costa Brava, *Barcelona*, containing the Maresme, the Barcelona beaches, and the Costa del Garraf, and *Tarragona,* which includes the Costa Dorada and the Ebro Delta. The data presented in this thesis is taken from buoys located within each of these provinces.

1.2.2 Economic Information

There are approximately 50 ports located along the coast of Catalonia (capitanes.com, 2009), and because of their high economic value to the area, protection of these ports is a top priority of port officials. For example, the port of Barcelona connects to over 825 ports worldwide. It contains two international container terminals, one terminal for multipurpose ships, seven international passenger terminals along with over twenty km of wharves and berths as well as employing 15000 people and 600 companies (PortdeBarcelona.es, 2010). The total traffic through the port in 2009 was just less than 43 million tons. Its importance is only emphasized by the fact that it is a leading European cruiser port and Mediterranean turnaround base.

Because of the Port of Barcelona's significant economic importance to the region of Catalonia, an expansion of the port was undertaken and completed in 2008. In order to adequately protect this additional area, two breakwaters were constructed called the South and East Breakwaters. The South breakwater is constructed in three sections. The first section is a 2000 meter long mound breakwater with an armor layer of parallelepiped blocks of concrete, each weighing 60 metric tons. The second section is 1700 meter vertical breakwater consisting of 47 prefabricated concrete caissons. The final section begins as a mound breakwater containing an armor layer of the same parallelepiped blocks, each weighing 40 metric tons, and ends with two prefabricated reinforced concrete caissons, all together stretching 1100 meters. The East Breakwater is a 2000 meter mound breakwater with a trunk of 50 metric ton parallelepiped blocks of concrete. The head is of the same blocks except each weighing 80 metric tons.

The second largest and most valuable port to the region is the Port of Tarragona, which traffics around 35 million tons of cargo volume each year (Porttarragona.es, 2009). The breakwaters surrounding the port's entrance are designed to protect against the following storm conditions in deep water: storms with a maximum wave height of seven meters and maximum wave length of 272 meters. The significant storm wave height in this region can reach up to 6.12 meters.

Considering the previous information regarding the storm conditions in the Mediterranean, it becomes quite clear why the understanding of storm patterns is necessary.

1.3 Risk associated with Mediterranean Storms

Because of its close proximity to the Mediterranean Sea, some of the main challenges facing coastal engineers in Catalonia is minimizing risks associated with flooding, wave damage, beach erosion, etc… This next section discusses briefly discusses the terms associated with defining and determining risk and then proceeds to explain specific risks associated with Catalonia's location along the Mediterranean.

1.3.1 Definition of Risk, Hazard, and Vulnerability

The definition of *risk* will vary amongst different disciplines, but in general, the UNISDR (United Nations International Strategy for Disaster Reduction) defines risk as the "combination of the probability of an event and its negative consequences" (UNISDR 2009). In particular, the idea of "natural risk" is of interest. Natural risk focuses on the "potential loss to the exposed subject or system" affected by an event and its concurrent consequence (Llasat, Llasat-Boteja, Lopez, 2009). It introduces two more concepts related to natural risk, namely, the concepts of *hazard* and *vulnerability.*

UNISDR (2009) defines hazard as a "dangerous phenomenon, substance, human activity or condition that may cause loss of life, injury or other health impact, property damage, loss of livelihoods and services, social and economic disruption, or environmental damage". Hazard is measured in terms of probability, which, in turn, is based on the frequency of occurrence of an event. Vulnerability is defines as "the characteristics and circumstances of a community, system, or asset that make is susceptible to the damaging effects of a hazard" (UNISDR 2009). Scientists of different disciplines have differing ideas on specific characteristics and circumstances, but generally, vulnerability is related to the capability of a system to respond to a hazard or protect itself from a damaging event. This implies that it is socially, politically, and economically dependent. The reliance of vulnerability on these three varying factors makes it difficult to quantify. Another aspect of vulnerability that adds to its complexity is that *indirect* consequences must be considered as well. Aspects of a system that cannot be defined with a monetary value, for example a wildlife habitat with a high ecological value, need also to be accounted for when discussing vulnerability. All these factors imply that great care should be taken when analyzing the effects of a hazard. In the case study conducted of two flood events that occurred along the Catalan coast in 2002, the assumption that the hazard map and the flood risk map are equivalent would lead scientists to incorrect conclusions concerning the impact of the events (Llasat, Llasat-Boteja, Lopez, 2009). This example can go to show how misleading a vague understanding of important concepts can lead to poor decisions-making.

The dissimilarities of opinion mentioned above bring about the idea of *risk perception*, which bases risk not on the actual occurrence of the event. The events are perceived as 'neutral' and the components of vulnerability then take over to define the natural risk. This brings about the idea of risk depending on *perception*, since the actual risk is based on the specific sensitivity to the event of the effected system. Therefore, the idea of natural risk is "relative" (Llasat, Llasat-Boteja, Lopez, 2009).

1.3.2 Specific Risk from Wave Storms in Catalonia

The risk associated with the presence of a twin storm must take into account an additional component of cost or damage, specifically, the cost or damage associated with the second storm. The second storm may play an important role in the total damages caused by the pair, since the first storm could have significantly weakened the means of defense so that the second storm is imposing on a much more vulnerable system. For example, if a storm has had enough time to erode a beach protecting a promenade, the promenade will then be exposed to the attack of the following storm. The same situation can occur to breakwaters protecting harbors that have experienced damage from the first storm and cannot adequately protect the assets of the port or harbor during the arrival of the second storm. This thesis will focus on how mitigating the damage induced by the second storm by applying a precautionary action between the storm pair can reduce the risk associated with wave storms.

Methodology

Determining a procedure to evaluate the factors discussed above will be the scope of this section. The first segment will outline the method in determining the presence of storms in the four areas of analysis, illustrating details about the buoy readings and specifying the criteria used to determining storm events. This will be followed by the implementation of the storm information into a probability model for the determination of significant variables. The section begins with a brief description of the model and tests of independence that have been applied to the data. After, definitions of useful parameters used for gauging the probability of twin storm events are discussed. This concludes the methodology used in reference to the probability of a twin occurrence. The next segment discusses the cost analysis performed on the twin event. Specifically, it discusses a way to decide when acting during the calm period in the middle of the twin event will prove to be beneficial. It begins by defining the variables utilized and their limits. Next, the application of these variables will be shown for determining associated risk, introducing a possible cost ratio and its criterion of usage. Closing the section of *Methodology* will be the possible areas of implementation.

2.1 Storms

Because of the varying characteristics of the Catalan coast, it can be safe to assume that the details of the storms are also different from one area to the next. It should be noted that this analysis solely looks at the wave heights. It does not take into account any other common characteristics of storms (surge, set-up).

2.1.1 Threshold Calculations

The first step in determining the number of storms is to set a threshold value which will act as the minimum wave height signifying a wave storm. Finding a threshold is done using the average excess method (Verhagen, d'Angremond, van Roode, F, 2009). All the readings from all the buoys are implemented into this analysis. The estimated thresholds are all the values taken above 100 cm in intervals of twenty (100, 120,...). The estimated thresholds are then subtracted from all the wave height values greater than this particular value. The average of the differences is calculated for each interval. Plotting these excesses gives curves as those seen in Figure 2. Next, a trend line is fitted to the data. If the data does, in fact, follow the predicted model, then the plots should follow the trend line. For the threshold values were this is concluded to be true, the model can be applied. This estimated threshold signifies a wave height value at which taking larger wave heights will not contribute to the analysis. The following figure displays the results:

Threshold Calculations

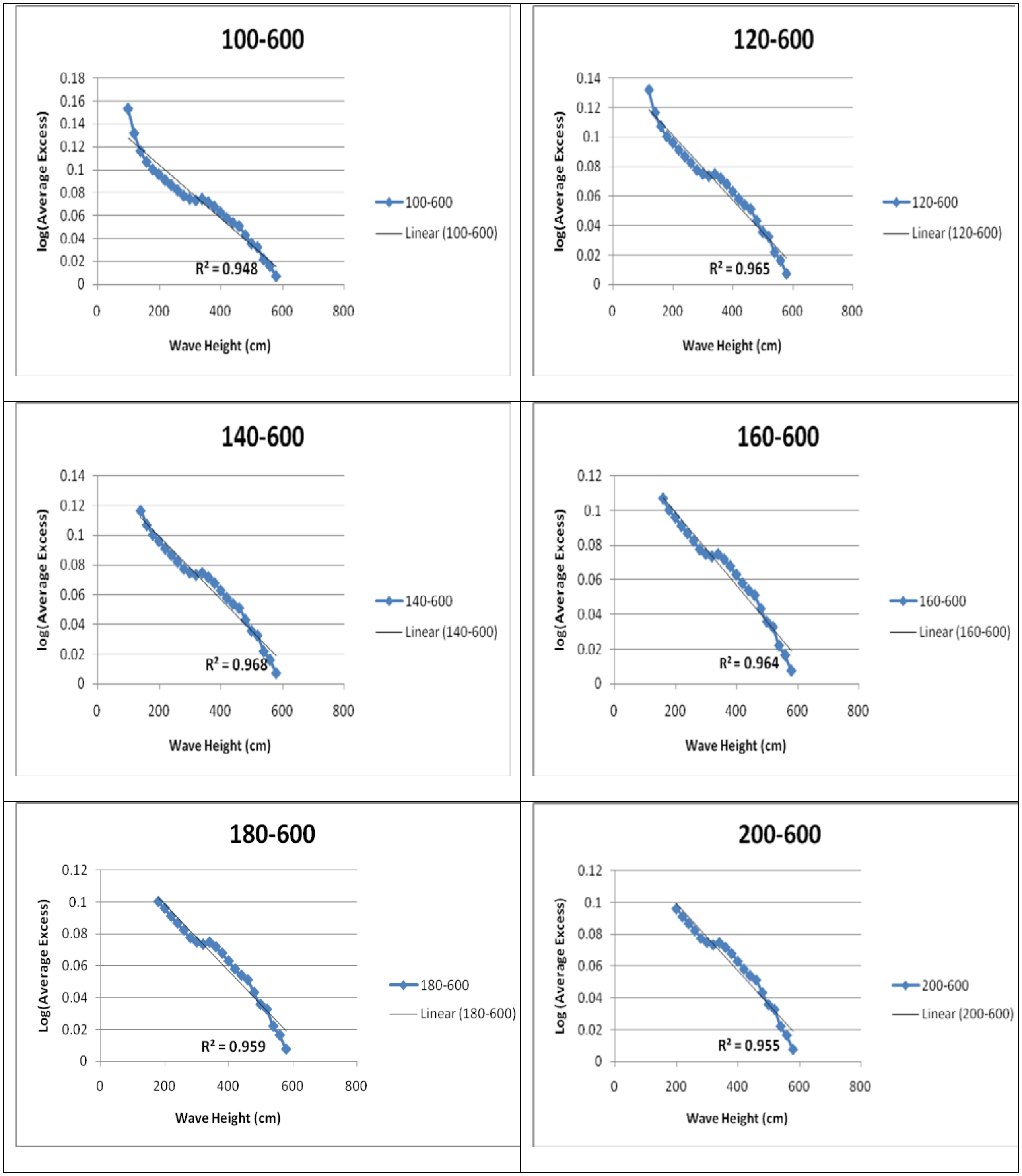

Figure 2: Threshold Calculations

As can be seen from the figure, the plot associated with a threshold of 140 cm is the closest representation of a line. However, in order to stay in the range suggested by the *Puertos del Estado* (Sanchez-Arcilla, Gonzalez-Marco, Bolanos, 2008), which advises a value between 1.5 and two meters, a threshold value of 160 cm is chosen for the analysis.

2.1.2 Filtering procedure

Once the threshold value is set, all heights below 160 cm are disregarded. The first step is to categorize the heights based on the days between each reading. Typically, wave heights that are separated by twelve or more hours of calm water are considered to be two autonomous storms (Jiménez, Sánchez, Valdemoro, Gracia and Nieto, 1997). However, in order to be absolutely sure that wave heights from the same storm are not considered two independent events, a difference of more than three days (72 hours) between readings is chosen to separate storms. Next, the number of buoy readings between each measurement is calculated. Based on the frequency of the readings, it could be decided which points present in the same storm. This frequency varies from buoy to buoy, as will be seen later; therefore the maximum value utilized varies as well. Then, a new criterion related to the storm duration is needed to separate these groups of buoy readings.

The next chosen requirement is that the wave height needs to be over 160 cm for a minimum of six consecutive hours to be considered a storm. This value is chosen based on the commonly accepted idea that six hours will provide enough time for typical storm-induced coastal processes to occur (Mendoza, 2008). This analysis is performed by taking the time difference between consecutive large wave heights in a group. Any group lasting less than six hours is eliminated so as to avoid freak waves being considered wave storms. Once each storm is identified, the maximum significant wave height is taken of all the heights recorded in this time period and defined as the storm wave height. The question then becomes whether each storm is one single event or a twin.

2.1.3 Definition of a Twin

There is no standard definition of a twin storm because still little is known about this phenomenon. The general consensus within the scientific community is that a twin storm is the occurrence of a storm within a certain period of a time following a previous storm. The specifications vary somewhat. This analysis labels a twin using the same criteria as a storm described above. In addition, a twin is based on the conclusion that the storm presents no less than six hours but no more than 72 hours after the initial storm. If significant wave heights are recorded within 6 hours, they are considered to be part of the previous event. Similarly, if more than 72 hours of calm water has passed, then the new heights represent a new event.

Unfortunately, the process of determining the occurrence of a twin storm is a visual one. Each storm (or in this case, each group of consecutive wave heights) has to be carefully examined in order to eliminate any chance of error. The criteria are as mentioned above: a storm must last for at least six hours (with a high level of confidence that this period was reached), as must its twin. In many cases, the buoy would malfunction and cease taking readings during the storm. It then becomes a matter of accurate extrapolation from the available data. In any situation where the data seemed questionable, for example a few hours pass with no recordings, all the heights from this group are disregarded. A storm is only labeled as a twin if the data shows without doubt that the criteria have been met.

Once a group has been decided as an individual event or as a storm with a twin, a zero or one was assigned to each, respectively. However, it is important to note that only the first storm is of significance in this analysis. That is to say, if a storm is labeled with a one, only the initial group of heights is used to define the storm based on storm wave height and storm duration. The 'twin' storm is not considered. The only exception to this is when the twin of a storm *has* a twin itself. This is then treated as a separate occurrence of a twin storm. This is the only incident when the data readings of the twin are used in the evaluation. In this particular situation, this happens rarely.

2.2 Specific Characteristics of the Data and the Data Area

This study uses information gathered by four buoys in order to determine the occurrence of a twin storm. The location of each buoy is represented in the following figure:

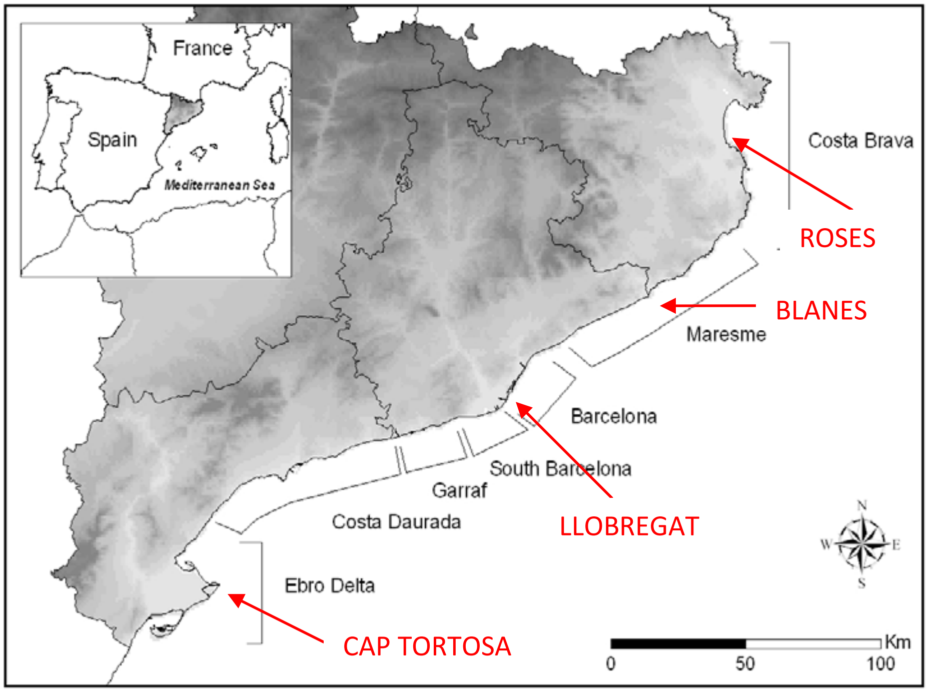

Figure 3: Buoy Location along the Catalan Coast

These buoys are all owned and operated by *Xarxa d' Instrumentació Oceanográfica i Meteorológica* (abbreviated *XIOM*), a segment of the *Generalitat de Catalunya.* The coordinate locations and measuring depths are:

Buoy	Latitude (North)	Longitude (East)	Depth (m)
Cap Tortosa	40 43.29	00 58.89	60
Llobregat	41 16.69	02 08.28	45
Blanes	41 38.81	02 48.93	74
Roses	42 10.79	03 11.99	46

Table 1: Buoy Position (Mendoza, 2008)

Details about the area and the buoys will be presented in the following discussion.

2.2.1 Cap Tortosa

The Cap Tortosa buoy is a directional Waverider buoy, meaning that it has the capacity to take directional data as well as scalar (wave height, period, etc…). However, this information is not utilized in this analysis. Only the wave height and the hour and day recorded are applicable. The data readings began in June 1990. During the beginning of the 90's, measurements were taken every three hours; the measurements, however, increased in frequency throughout the 90's. Also, gaps were present in the data readings. Almost every year had at least one month and up to three months (1995, 1998) of missing measurements. This problem diminished in the 2000's with only week-long gaps in measurements. Also, by this time, the readings began to be taken every hour, on the hour. It is worth mentioning that the gaps in the information do not seems to follow any particular pattern, such as a seasonal dependence or occurring in a year with unusually strong Mediterranean storms. Therefore, it can be concluded that it is quite likely that strong storm activity is not solely responsible for a disruption in measurements.

2.2.2 Blanes Characteristics

The recordings in Blanes are of significantly lesser quality than those in Cap Tortosa. The buoy is a similar Waverider with the exception of being scalar, not directional. The measurements begin in April of 1984 and continue until June of 1997. At best, the recordings are taken every three hours. Throughout the 80's, however, they are taken every four hours, which makes determining the occurrence of a twin based on the criteria previously mentioned impossible during this time period. In actuality, only one twin can confidentially be documented, in the year 1988. Because of this limitation, only the data from the 90's should be applied, which unfortunately limits the accuracy to which the model can fit the data.

2.2.3 Roses characteristics

The buoy at Roses is a Waverider scalar buoy located in the Gulf of Roses in the Costa Brava. The sampling rate is every three hours. The readings started in September of 1992 and continued steadily until September of 1996. The buoy was not operational until the following February in 1997, but only recorded briefly before discontinuing measurements between July 1997 and May 2000. The readings then continued until May 2005 at one hour intervals with only maximum month-long breaks between recordings.

Besides the decreased period of measurements and the increased inconsistencies in the measurements, the information taken at Roses has yet another disadvantage. Its location is inside a harbor that is protected with coastal structures. This implies that diffraction of the waves around the structures cannot be ignored and ultimately results in decreased accuracy of the measurements. In spite of these short-comings, the data is still analyzed and outlined in this thesis.

2.2.4 Llobregat Characteristics

The Llobregat buoy is a Waverider which preformed as a scalar until February of 2004 and as a directional buoy then after. The readings started in July of 1984. For the next six years, measurements were taken in four hour intervals, until the 90's when they began alternating between three and four hours. The measurements abruptly stopped at the end of '96 and did not start again for three years. In the 2000's, the readings began to be taken every hour. The measurements utilized in this analysis are taken until the end of 2007.

It should be noted that accuracy of twin predictions is highly sensitive to the measurement intervals. Therefore, the 2000's give a much more accurate depiction of the occurrence of twins because the data can be more closely scrutinize in order to confirm that the criteria is met. In addition, duration of the storm is dependent on the frequency of the measurements as well. During the 80's, the recordings occurred every four hours for the buoys of Llobregat and Blanes. For this reason, the criterion of six hour storm duration and six hour breaks are increased to at least three consecutive readings, or 12 hours. This implies that for a storm to be considered a storm during the 80's, there must be at least three measurements, or a 12 hour duration. The same is for a storm to be considered as having a twin (three consecutive readings of calm water). This limitation also limits the accuracy of the duration variable because each duration term is represented as a multiple of four, which can grossly overestimate the actual duration of the storm.

2.3 Probability Model Information

The next step in this twin storm evaluation is determining the best fit probability model to the information provided. In general, a linear model is used to define the dependent response variable by using the mean parameters. The mean parameters are, in turn, described by the linear combination of the independent co-variables and the spread parameters. This particular assessment calls for a model that accurately relates the dependency of a twin storm event on the proposed variables. The following section presents a brief introduction to the chosen model, tests, and parameters utilized in order to characterize twin storms.

2.3.1 Generalized Linear Models

In order to determine whether there exists a dependence of the occurrence of a twin storm on the variables mentioned in the *Storms* subsection of *Methodology*, a generalized linear model is utilized. This model is chosen based on its extended range of probability distributions. The linear model assumes the conditional distribution on the co-variables is normal whereas the Generalized Linear Model (abbreviated *GLM*) allows for other distributions in the exponential family to be fitted to the data, for example the Poisson, Gamma, or Binomial distribution.

When evaluating binary data, the two most important GLM's are the *logistic regression model* and the *log-linear* model (Agresti, 2007). Specifically, the logistic regression model is applied to data that contains a mixture of continuous and categorical variables while the log-linear model is used with only categorical variables. Both require the dependent variable to be dichotomous. In this analysis, the dependent variable is the response. It can also be called a Bernoulli variable since the occurrence of a twin can be labeled as a success or failure. The number of storm events in this data set can be referred to as Bernoulli trails each resulting in a success (twin) or failure (no twin). As mentioned before, the presence of a twin will take the value of 1, while a storm occurring alone will be given a value of 0. Also, the presence of a binary response variable alludes to the use of the binomial distribution.

There are three components of a *GLM*, the random component, the systematic component, and the link function. The random component is the response variable. The systematic component is the explanatory or predictor variables. As the name insinuates, they are used to explain or predict the response variable in mean behavior; each is assumed to exhibit a linear influence on the response variable. The link function serves as a link between the systematic and random components. It expresses the expected value or mean of the probability distribution. In a log-linear model, the link function is:

$$\log(\pi(x)) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \tag{1}$$

As mentioned before, the log-link model is appropriate for count data where the values are non-negative. The log-link function will present the log of the mean of the data.

When the mean or expected value is desired as a probability, the *logit* link function is more appropriate. The logit takes the log of the odds ratio and can be expressed as:

$$\mathit{logit}[\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \tag{2}$$

The logit is a form of defining the log odds of a particular outcome. The corresponding *GLM* associated with the logit link function is the logistic regression model. It should be noted that the use of logistic regression does not describe the distribution of the explanatory variables but only that of the dependent or response variable. The odds can be considered unfavorable for the occurrence of the event if the logit takes a negative value, and the event is considered likely if the odds are positive. In this equation, α is the constant, or intercept, of the logit. It represents the value of the log odds when the measure of the contribution of all the independent variables is zero. The β's are the predictor coefficients. Essentially, each coefficient will give you the log odds ratio of each particular predictor variable (Agresti, 2002). The sign of the predictor coefficients will determine the level of dependency of the response on each predictor. A positive β will imply that the variable increases the odds of occurrence while a negative coefficient signifies that the predictor decreases the odds. Further interpretation based on the value of the coefficients is somewhat more difficult to decipher.

2.3.2 Tests of Independence

When fitting any model to a set of data, it is imperative to know on which explanatory variables the response variable is dependent. In other words, the independence between the variables must be tested to determine the relationship between them. In this analysis, this is done by dividing the storms with a positive response to a twin into a separate group from the storms with a negative response to a twin. The following discussion outlines the tests used in this analysis.

The descriptive statistics of the two groups will be compared to each other using the ANOVA and Chi-square tests, which are presented in the following section. Both methods make use of the null hypothesis which serves as a standard for comparison. It states that each variable is considered independent from the other variables. A likely null hypothesis indicates that the descriptive statistics of the groups are essentially the same and that no specific variable initiates any variance between the two groups. An unlikely null hypothesis implies that an explanatory variable or set of variables can affect the outcome of the trail, making one outcome more probable than the other. To measure the likeliness of the null hypothesis, a parameter called the *p-value* is used.

2.3.1.1 ANOVA: Analysis of Variance

The first section of this discussion describes the ANOVA, or *Analysis of* Variance. However, the ANOVA description is only meant to provide the reader with a general idea of the principles used when conducting the analysis. This ANOVA cannot be applied to this particular study due to that fact that the data cannot be fit to a linear model because it cannot be assumed that it will follow a normal distribution. Instead, a variation of the ANOVA, called the *Analysis of Deviance* is utilized. A discussion of the Analysis of Deviance will be presented following the information about the ANOVA.

Generally, ANOVA is applied in order to compare the means of two populations which are assumed to have a normal response. In this case, the two populations refer to the *Twin?* or *No Twin?* groups. ANOVA utilizes the *F statistic* which is a ratio measure of the variance resulting from the difference between the groups to the error variance. A large F statistic implies that there is a large difference in variance between the two groups. Therefore, if the null hypothesis refers to independence between the groups, a large value of the F statistic would result in a rejection of the null hypothesis, namely, that the two groups cannot be considered independent of each other. A small value of the F statistic (usually less than 1) indicates that the explanatory variable does not have an effect on the response variable, hence validating the null hypothesis of independence.

When there are multiple factors that can affect the means of both populations, ANOVA can be applied in order to determine which factor contributes the most influence to the variation in the means of the two groups. That is to say, if a factor has a large F statistic, it is highly responsible for the difference in means of the populations. If the F statistic is low, the factor is not responsible for any difference between the two groups.

The next important aspect of ANOVA application is the effects that each factor has on the dependent variable. These can either be classified as main or interaction effects. Main effects refer to the direct influence that one independent variable can have on the dependent variable. If the independent variable is a factor that is split into multiple levels, as is the case in this paper, the main effect simply defines the comparison of the mean at each level to the mean at every other level of that particular factor. Interaction effects refer to the influence of the joint action of two independent variables on the dependent variable. Because each variable must be compared to all the others, as the number of independent variables increases, there will be an increase in interaction terms between the variables. Therefore, ANOVA designs are classified based on the number of independent variables they contain. A one-way ANOVA has only one variable that produces one single main effect, hence only one term. When more than one variable is present, there may be an interaction term, when there are more than two variable, there may be interaction terms and two-way interaction terms, which will account for all three variable influencing each other and having a resulting influence on the dependent variable. A four-way analysis (with four terms) will also have a three-way interaction term. This can quickly complicate an ANOVA design, making it slightly impractical for models with many variables.

The Analysis of Deviance differs due to the fact that it contains a dispersion parameter for the fitted family of distribution specified (Hastie, Pregibon, 1992). An implementation of this in "R" will display a table with the following variables:

Degree of Freedom	Degrees of freedom associated with each source
Deviance	Deviance for each source term
Residual Degree of Freedom	Number of data points available minus the number of parameters
Residual Deviance	Residuals are typically used to quantify a lack of fit of a model. This output defines a residual that incorporates a contribution from each observation to the deviance statistic (McCullagh, Nelder, 1989).
P=(>|Chi|)	Chi-square test specified for the binomial distribution: Used to compare the reduction of deviance in each source to the residuals when the dispersion is known. This reduction will follow a Chi-square distribution.

The table above makes reference to the *Chi-square test of independence,* another method used to check the independence between the variables. In order to understand the mechanism used in this test, the idea of contingency tables should be farther introduced.

2.3.1.2 Contingency Tables

In situations where all variables are categorical, the information is organized into a *contingency table*. These tables will be seen later in this analysis when comparing the occurrence of twins within the different seasons (for example, in **Table 11**). Contingency tables contain the frequency of occurrence of each response variable corresponding to each categorical predictor. Three probabilities can be defined using these tables: the *joint, marginal,* or *conditional* probability (Agresti, 2007). The joint probability is defined by probability of each particular cell in the table by dividing each cell frequency by the total number of trials. The marginal probability is used to define the probability of occurrence of a single variable. In the table, it is represented as the summation over each column or row. The idea of marginal probability is important in defining variable independence. For example, if a twin is decided to be independent of the season, then the occurrence of a twin in spring will be equal to the product of the marginal probabilities of the all the twins and all the storms in spring. Numerically:

Probability of a Twin in Spring: $$\pi_{Twin(Spring)} = \pi_{Twin} * \pi_{Storm(Spring)} \qquad (3)$$

If the null hypothesis is found to be likely, indicating independence between each variable in the table, the joint probabilities become the marginal probability.

Since this analysis contains explanatory variable and response variables, conditional probabilities are better suited for this analysis. Conditional probabilities refer to the probability of one response variable given the level of the predictor variable. This is useful in defining the conditional distribution of the response variable.

2.3.2.3 Chi-square test

With this understanding of the structure of contingency tables, the description of the Chi-square test can continue. Specifically, the *Pearson Chi-square test* is applied in this analysis, which is represented by:

$$\chi^2 = \sum \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} \qquad (4)$$

where *i* and *j* represent cell locations in a contingency table. This statistic is referred to as the *Pearson Chi-square statistic.* The magnitude of its value is used to determine how closely the data fits the null hypothesis.

This method is usually used with discrete data made up of qualitative variables. One specific type of qualitative variable is the *nominal* variable, which refers to variables that are categorized into groups or classes (Gaur, 2007). Descriptive statistics do not offer much useful insight for nominal data; frequencies are used instead to describe the data. These frequencies are then presented in contingency tables. The Chi-square test compares the observed cell frequencies with the expected cell frequencies present in the contingency table, which is referred to as the *significant level.* The expected frequency is calculated by multiplying the total of the row by the total of the column in which the cell is located in the contingency table and then dividing the value by the total of all the rows and columns (Gaur, 2007). If the expected frequency is less than the observed frequency, meaning their ratio is less than one, it can be concluded that there exists a correlation between the variables; therefore, the null hypothesis of independence should be rejected. The p-value is essentially just a measure of the likelihood of the null hypothesis.

2.3.3 Odds Ratio and Relative Risk

Another useful parameter that can be derived from a contingency table is the *odds ratio*. For a Bernoulli trial, the odds ratio is define as the odds of success. This can be written as:

$$odds = \pi_{success} / \pi_{failure} = \pi_{success} / 1 - \pi_{success} \tag{5}$$

The odds ratio is used to gauge the likelihood of a success. The larger the ratio, the more probable it is to obtain a success. The same goes for the opposite, a ratio less than one will indicate a failure more likely than a success. The only boundary on the odds ratio is that it will always be a number greater than zero.

In contingency tables, the odds ratio can be applied among rows and columns. Using the probability of success of two separate categories and comparing them in an odds ratio will result in a ratio signifying which success is more likely in which category. The odds ratio is also used to determine the level of independence between the rows (or categories) in the table.

Large differences in likelihood will yield large odds ratios, indicating that the response variable is dependent on the category. Values closer to one imply stronger association and an odds ratio of one denotes complete independence (Agresti, 2007). Using the joint probabilities, the odds ratio for a response variable of success and failure is:

$$OR = \frac{\pi_{1,success} * \pi_{2,\, failure}}{\pi_{1,\, failure} * \pi_{2,success}} \tag{6}$$

where 1 and 2 denote the categories being compared. If these values were taken as a 2X2 independent table, the odds ratio can be thought of as the ratio of the products of diagonal components. Therefore, it is also known as the *cross-product ratio*.

In order for the odds ratio to be meaningful in the storm analysis in Catalonia, it must somehow be correlated to the relative risk associated with the occurrence of a twin. Conveniently, the odds ratio can be evaluated as a *relative risk.* Relative risk is the ratio of proportions. It should not be confused with 'risk' described above (probability * vulnerability), because it is only a ratio of probabilities. It will be seen later that this parameter can be applied to the risk equation when making a comparison to risk scenarios. The odds ratio relates to the relative risk through:

$$RR = OR * \frac{1-\pi_{1,success}}{1-\pi_{2,success}} \tag{7}$$

2.4 Chosen Factors Fitted to the Probability Model

Utilizing the data provided by the buoys, the dependence of a twin is tested against intensity (storm wave height), duration of the first storm, and date; all of which are continuous variables. Each event has a value for each variable that is independent of all other events in the data set. Once a storm is detected, the intensity, duration, and date of that storm are recorded as described above. In addition, the occurrence of a twin is assigned to each event with either a one or a zero.

The first continuous variables of interest are the storm wave height and the duration. The storm wave height is calculated to be the maximum significant wave height measured during the storm. Since the readings were exact up to the centimeter, this variable is very precise. However, since the measurements were taken at different intervals during different decades, one particular measure may represent the significant wave height of either one, three, or four hours of measurements. The duration is a bit more estimated since it depends on the interval of measurement more directly. Intervals of four hours will provide durations that are multiples of four; the same can be said for three hour intervals.

The next continuous variable of interest is the time of year of the storm. To account for this factor, the date is taken as a periodic value. A sinusoidal function with a period based on the number of days in one year (*T= 365.25)* is used to describe the days in which the storms occurred. This variable now has two components, which may either, both, or neither influence the occurrence of a twin. Each interaction of the components will be encompassed in the shape of an ellipse. The ellipse can be used as a tool along with the ANOVA and Chi-square test to determine the dependence of the date components on the occurrence of a storm. If the point at which the component equals zero is contained in the ellipse, that component has no influence on the storm event. If the intersection of the sine and cosine term both equaling zero is within the ellipse, this is an indication of neither component influencing the event, therefore concluding that the storm occurrence is independent of the date.

If the date proves not to be significant as a periodic function, a new way to represent the time of year is applied. Each storm is categorized into the season in which it occurred. This implies that the season's variable is not continuous, but categorical. Therefore, when this variable is implemented into the generalized linear model, it is not done so as a predictor, but as a modifier for the wave height and duration variable. This will produce not only main effects but also interaction effects that may prove to be significant.

2.5 Cost and Risk Analysis

Once the process of probability evaluation is complete, the next step deals with evaluating damages associated with twin events. This section will begin by introducing the applied cost parameters, specifically their definitions and boundaries. The costs will then be parameterized into a single unit-less parameter, SCR_T, which will be used in the following risk analysis. A table of values will be presented of how to quantify the SCR_T based on the parameters on which it depends. The risk analysis will display a relationship between the probability of a twin and this derived cost factor, SCR_T.

2.5.1 Cost

Once the probability of the hazard has been determined, attention must be given to the consequences of a twin storm event, such as structural damage or loss of beach width. This analysis is done in terms of cost. However, it is kept general so that other quantifiers of damage can be applied.

The first step is to determine the costs associated with the presence of a twin storm event. The following variables are chosen to represent this occurrence:

Cost Variable	Description	Value
C_0:	Cost of the first storm	= Constant
C_1:	Cost of the twin storm if no prior action is taken	= C_1
C_2:	Cost of the twin storm if action is taken.	= βC_1

Table 2: Definition of Cost Parameters

C_0, the cost of damage associated with the first storm, is an inevitable cost. It is assumed a constant in this analysis due to the fact that alleviating it is not within the scope of this thesis. The second cost, C_1, refers to the cost of damage caused by the second storm when nothing is done to mitigate the event. This implies that the *total* cost of the twin storm is the sum of these two costs:

$$C_{total} = C_0 + C_1 \tag{8}$$

The third cost, C_2, refers to the cost of the second storm if action is taken to prevent damage from occurring. When conducting this analysis using monetary values, the cost of the action can be included in this variable. This means that it will be the sum of the cost of action and the cost of the damage endured after the second storm has passed. However, this is not possible if the consequence is being evaluated using some other scale (how would one relate applying sand bags along a shorefront with the beach width or sand lost during the second storm of the twin set?). In **Table 2**, C_2 is related to C_1 by a variable β. This variable is defined as follows:

β:	Ratio of the cost of second storm with prior action taken to cost of second storm with no action	$= C_2 / C_1$; $0 < \beta < 1$

β defines *how much* prior action can reduce costs associated with the second storm. This implies that having a value greater than one would indicate acting between storms would produce costs greater than not acting between storms, deeming it impractical to act. This is a concern when evaluating the practicality in taking preemptive measures. Because of this, the range of β is the first criterion in this cost analysis. β = 0 is the idealistic case that total damage from the second storm can be prevented, and β =1 refers to the cost of acting being the same as the cost of not acting, and therefore not producing any benefits.

The total cost of any of these three storm events is then parameterized in order to insert it into a risk analysis later in this section. This is done by dividing the cost of each storm scenario by the cost of the first storm, since it is the cost that will be endured in any storm event. This yields:

SC_{NT}:	Specific Cost of Storm with no Twin	= 1
SC_{TNA}:	Specific Cost of Storm with Twin and No Action	$= \left(1 + C_1 / C_0\right) = (1+\rho)$
SC_{TA}:	Specific Cost of Storm with Twin and Action taken	$= \left(1 + \beta C_1 / C_0\right) = (1+\beta\rho)$
SCR_T:	Specific Cost Ratio of a Storm with a Twin of Not taking Action to taking action	$= \dfrac{(1+\rho)}{(1+\beta\rho)}$; $1 < SCR_T$

Table 3: Parameterized Costs

This introduces two more parameters into the analysis:

ρ :	Ratio of cost of second storm with no prior action to the cost of the first storm	$= C_1 / C_0$; $0 < \rho$
N:	Cost product	$= \beta * \rho$

Table 4: Parameter Definition

ρ, a parameter relating the cost of the second storm to the first, can also be informative about the level of damage that can be expected. As mentioned before, it only makes sense to act between storms if the damage induced by the second storm is large, which can very well be the case when a system becomes exposed due to the first storm and then is immediately hit by a second storm. The lower boundary of this parameter is zero, since a value of zero means that there is no second storm, and the storm event consists of only a single storm. ρ has no upper boundary, which indicates that the damage caused by the second storm can be significantly higher than the damage from the first on the vulnerable system.

The next parameter is the product of $\beta * \rho$, namely, *N*. It is essentially the ratio of the cost of the second storm with previous action to the cost of the first storm. It accounts for both the size of the damage induced by the occurrence of the second storm and the amount that acting will decrease the damage. Note that *N* will always be less than ρ since β is always less than one.

The last parameterized cost presented in the **Table 4** is the *Specific Cost Ratio of a Twin*, SCR_T. This value results from dividing the specific cost of not acting between twins and acting between twins. It is always greater than one. Its usefulness will emerge when determining the risk associated with the occurrence of twin storms.

2.5.2 Risk

In the introduction, the idea of risk was defined as *probability * consequence*. This analysis focuses on the risk associated with the presence of a twin storm. Therefore, the following two specific risks must be related:

SR_{TNA} :	Specific Risk of Storm with a Twin and No Action	$P_T * (1 + \rho)$
SR_{TA} :	Specific Risk of Storm with Twin and Action	$P_T * (1 + N)$

Table 5: Specific Risk of a Twin Storm

The first part of the methodology described how to determine the probability of the twin. Now, applying that probability and the specific cost will lead to an analysis of the specific risk associated with twin storm events. Evaluating the specific risk will hopefully give some insight into when action is desirable:

$$R = P_T SCR_T \quad = \quad P_T * \frac{(1+\rho)}{(1+N)} = P_T * \frac{(1+\rho)}{(1+\beta\rho)} \tag{9}$$

The above risk should help to determine the necessity to perform some anticipatory measures in expectation of a twin storm. In order to do this, a value of SCR_T must be defined so that the probability of a twin can be inserted into the equation.

In order to calculate a legitimate value for the risk, this analysis will only make sense if SCR_T > *1*. This will always occur if β is less than one (acting between twin events does mitigate the damages indured). If this criterion is not met, and SCR_T is less than one, then it could be concluded that it is cheaper to not act than to act. Therefore, this analysis necessitates finding suitable values of β, ρ, and *N.* In order for acting between a storm and its twin to make sense, the following should be true. Since β is a factor to define the similarity in costs of acting before a twin and not acting, it should be as low as possible. ρ quantifies how much more destructive a sequential storm will be in comparison to the first, therefore should be as high as possible for acting to be practical. High values of ρ are possible if the increased vulnerability of a beach, (i.e. exposed promenade) are taken into account.

The following plot shows the SCR_T vs. the P_T for a risk of 1 and a risk of 2.

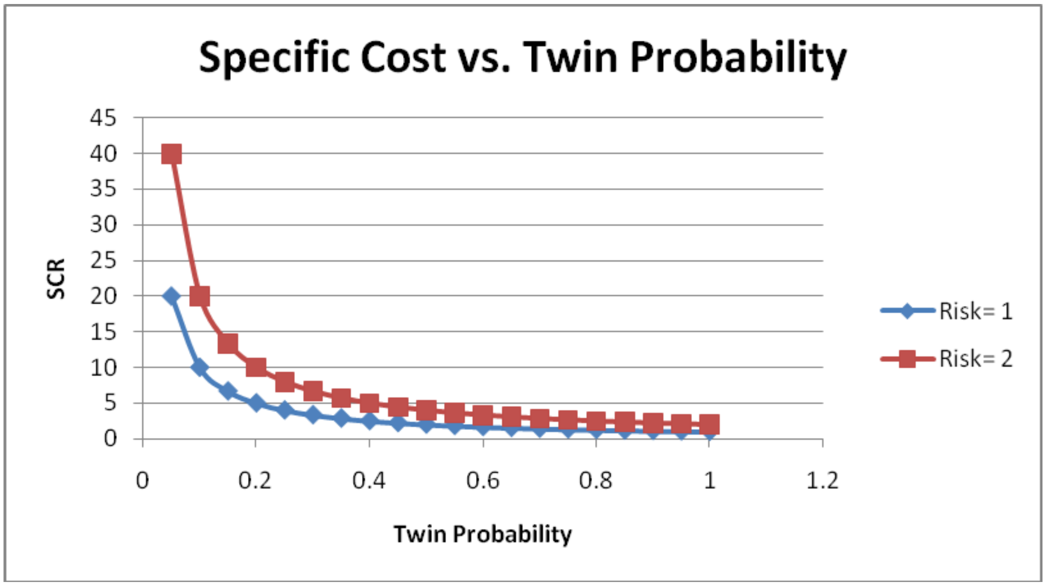

Figure 4: Twin Probability and Cost

If the values of β and ρ are know, the graph can be followed to find an acceptable probability. If, however, β and ρ are the unknown parameters, the same can be done to find their value. Taking a known probability and risk of a twin and finding the SCR_T associate with it, the values can be deferred from it.

Another benefit of examining the plot will result from observing when the curve begins to flatten. This point indicates that as probability increase, the SCR_T does not change significantly; therefore, the difference between acting and not acting is minimal. Taking this value of SCR_T and looking at the following table will give a minimum and maximum value for β and ρ. For example, the highlighted cells correspond with a SCR_T value larger than 2.9.

p/B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.2	1.176471	1.153846	1.132075	1.111111	1.090909	1.071429	1.052632	1.034483	1.016949
0.4	1.346154	1.296296	1.25	1.206897	1.166667	1.129032	1.09375	1.060606	1.029412
0.6	1.509434	1.428571	1.355932	1.290323	1.230769	1.176471	1.126761	1.081081	1.038961
0.8	1.666667	1.551724	1.451613	1.363636	1.285714	1.216216	1.153846	1.097561	1.046512
1	1.818182	1.666667	1.538462	1.428571	1.333333	1.25	1.176471	1.111111	1.052632
1.2	1.964286	1.774194	1.617647	1.486486	1.375	1.27907	1.195652	1.122449	1.057692
1.4	2.105263	1.875	1.690141	1.538462	1.411765	1.304348	1.212121	1.132075	1.061947
1.6	2.241379	1.969697	1.756757	1.585366	1.444444	1.326531	1.226415	1.140351	1.065574
1.8	2.372881	2.058824	1.818182	1.627907	1.473684	1.346154	1.238938	1.147541	1.068702
2	2.5	2.142857	1.875	1.666667	1.5	1.363636	1.25	1.153846	1.071429
2.2	2.622951	2.222222	1.927711	1.702128	1.52381	1.37931	1.259843	1.15942	1.073826
2.4	2.741935	2.297297	1.976744	1.734694	1.545455	1.393443	1.268657	1.164384	1.075949
2.6	2.857143	2.368421	2.022472	1.764706	1.565217	1.40625	1.276596	1.168831	1.077844
2.8	2.96875	2.435897	2.065217	1.792453	1.583333	1.41791	1.283784	1.17284	1.079545
3	3.076923	2.5	2.105263	1.818182	1.6	1.428571	1.290323	1.176471	1.081081
3.2	3.181818	2.560976	2.142857	1.842105	1.615385	1.438356	1.296296	1.179775	1.082474
3.4	3.283582	2.619048	2.178218	1.864407	1.62963	1.447368	1.301775	1.182796	1.083744
3.6	3.382353	2.674419	2.211538	1.885246	1.642857	1.455696	1.306818	1.185567	1.084906
3.8	3.478261	2.727273	2.242991	1.904762	1.655172	1.463415	1.311475	1.188119	1.085973
4	3.571429	2.777778	2.272727	1.923077	1.666667	1.470588	1.315789	1.190476	1.086957
4.2	3.661972	2.826087	2.300885	1.940299	1.677419	1.477273	1.319797	1.192661	1.087866
4.4	3.75	2.87234	2.327586	1.956522	1.6875	1.483516	1.323529	1.19469	1.08871
4.6	3.835616	2.916667	2.352941	1.971831	1.69697	1.489362	1.327014	1.196581	1.089494
4.8	3.918919	2.959184	2.377049	1.986301	1.705882	1.494845	1.330275	1.198347	1.090226
5	4	3	2.4	2	1.714286	1.5	1.333333	1.2	1.090909

Table 6: SCR_T **Values**

In other words, if the SCR_T value is estimated to be 2.9, then a twin event probability of 50% would result in an *R* of 1.45. However, if the probability drops to 33%, the *R* then drops below one to 0.957.

Since the value of ρ can be infinite, this analysis only incorporates values less than or equal to ten. This corresponds to a SCR_T of approximately 5.5.

Results

The following section will outline the results achieved from applying the above procedure for the four study areas, including the statistical descriptions and both the results of the *GLM* fitting and the tests of independence between the variables. In addition, the odds ratio will be applied to the relevant areas. Finally, the application of the cost and risk analysis to the probabilities found in Cap Tortosa and Llobregat will be presented.

3.1 Cap Tortosa

The first area examined is the Cap Tortosa because the data is of significantly better quality than the other buoys. Also, the most overall storms, and consequently twins, can be found in this area. In addition, the data readings are the most consistent of all the areas. The following tables show the descriptive statistics of the Cap Tortosa buoy.

Storm Wave Height			
Statistical Description	All	Twin	No Twin
No. of storms	216	78	138
Median	225	222	228
Standard Deviation	58	65	55
Average	241	241	242
Mode	214	216	260
Range	378	376	286
Skew	2	2	1

Storm Duration			
Statistical Description	All	Twin	No Twin
No. of storms	216	78	138
Median	16	15	16
Standard Deviation	17	16	18
Average	21	20	22
Mode	6	9	6
Range	121	84	119
Skew	2	2	3

Table 7: Statistics for Tortosa Storms

From the first table, the average height of all the storms, the storms with twins, and the storms with no twins are within a few centimeters of each other. The median storm wave height is within three centimeters of each other for all three categories. This implies that wave height variation between the groups is minimum, which is the first indication that storm wave height may not affect the occurrence of a twin. The same can be said for the average duration of the storms, each within one hour of the other. The conditional density plots illustrate this point even more:

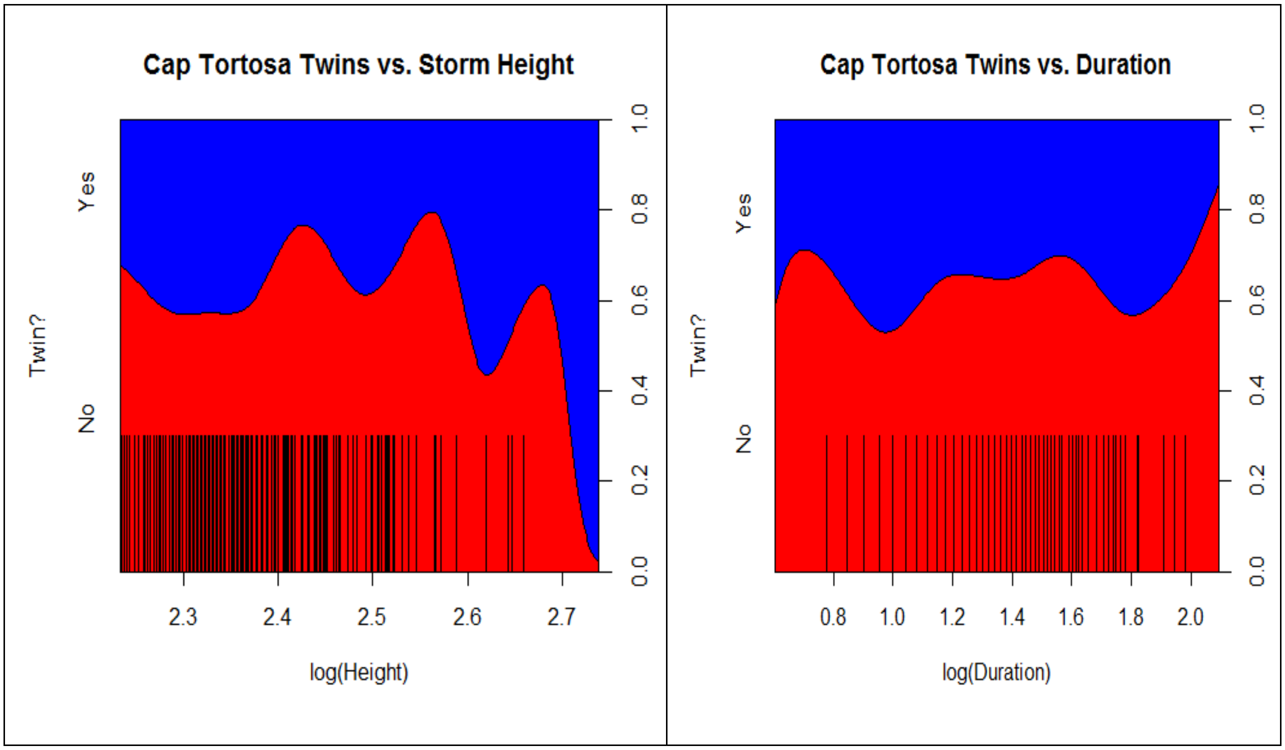

Figure 5: Conditional Density

In the wave height density plot, there does appear some variation, for example, the deepest trough at Hs between four and five meters (2.6 < logHs< 2.7) indicate that 60% of the storms are twins. However, on closer examination, it can be seem that only four storms fall into this range, containing two twins. Only one storm height is higher than five meters and is also a twin. The duration density plot reveals the same lack of correlation between the length of the storm and the twin. Therefore, it can be presumed that only date or season may influence the occurrence of a twin storm.

In order to test this hypothesis, the data was introduced into "R" and fit to a logistic regression model. As mentioned before, the dates are treated as periodic random variables, each day containing a sine and cosine component. The applied null hypothesis states that all four variables, namely: storm wave height, duration, sine (Day), and cosine (Day), are independent of each other and independent of the response variable (twin or no twin?). The formula and results provided by "R" can be found in Appendix F.

Both an Analysis of Deviance and Chi-square test were applied to the model to check if the null hypothesis did correctly predict the independence of the predictor variables against the response variable. The following table illustrates the obtained results:

	Degree of Freedom	Deviance	Residual Degree of Freedom	Residual Deviance	P=(>|Chi|)
NULL			215	282.55	
logHs	1	0.11876	214	282.43	0.7304
logD	1	0.40744	213	282.02	0.5233
sinDate	1	2.19073	212	279.83	0.1388
cosDate	1	0.7823	211	279.05	0.3764

Table 8: ANOVA and Chi-square test for Cap Tortosa Data

Observing the Chi-square results, all of the values are above the significant value of 0.05 or 0.10. As mentioned above, the test must result in a value that is less than a significant level of the ratio relating the observed frequency to the expected frequency of the variable; in this case, the significant levels are five and ten percent. As can be seen, none of the p-values fall below 0.10 or 0.05; therefore, the response is not dependent on any of the variables.

Since the assumption of a periodic-date influence showed not to be the case, a new way to represent the time of year becomes necessary. Dividing the storms into seasons provided a new categorical variable. The descriptive statistics for the seasonal storms in Cap Tortosa are shown below:

Seasonal Storm Wave Height					
Statistical Description	All	Spring	Summer	Fall	Winter
No. of storms	216	52	6	87	71
No. of Twins	78	24	0	30	24
No. of Not Twins	138	28	6	57	47
Median	225	219	225	229	225
Standard Deviation	58	61	34	65	49
Average	241	240	228	244	239
Mode	214	194	-	183	225
Range	378	265	78	378	196
Skew	2	2	0	2	1

Seasonal Storm Duration					
Statistical Description	All	Spring	Summer	Fall	Winter
No. of storms	216	52	6	87	71
No. of Twins	78	24	0	30	24
No. of Not Twins	138	28	6	57	47
Median	16	15	16	15	18
Standard Deviation	17	15	10	17	19
Average	21	19	19	20	23
Mode	6	6	-	9	10
Range	121	62	26	119	90
Skew	2	2	1	3	2

Table 9: Statistical Description for the Seasonal Tortosa Storms

The first notable observation from the above information is that no twins occur in summer. Overall, very few storms occur in summer. Therefore, the following discussions will not include this season but focus on the others.

Running an ANOVA and Chi-square test yields the following results:

	Degree of Freedom	Deviance	Residual Degree of Freedom	Residual Deviance	P=(>|Chi|)
NULL			215	282.55	
logHs	1	0.1188	214	282.43	0.7304
logD	1	0.189	210	274.37	0.66376
Season	3	1.8729	211	274.56	0.04871
Season:logHs	3	2.5777	207	271.79	0.46141
Season:logD	3	4.1288	204	267.67	0.24789

Table 10: ANOVA and Chi-square including seasonal information

The p-value of the ‘season’ variable (0.049) is less than the pre-selected significant level of 0.05; it can therefore be concluded that the NULL hypothesis of independence can be rejected. In other words, there is a relation between the season and the occurrence of a twin storm. This relation can be seen in the following plot:

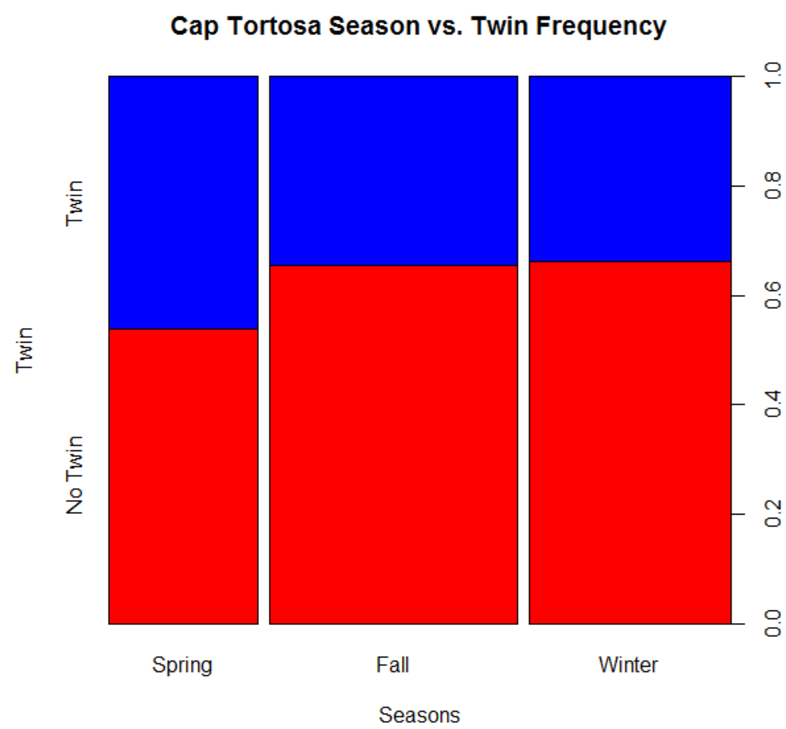

Figure 6: Tortosa plot of Twin vs. Seasons

From the mosaic, there seems to be not much difference in the occurrence of a twin between the seasons. However, for a closer examination, the odds ratio will be calculated for a twin storm in each season as well as the odds within the seasons.

The contingency table for counts is as follows:

Season	No Twin	Twin	**Total**
Spring	28	24	52
Fall	57	30	87
Winter	47	24	71
Total	132	78	210

Table 11: Contingency Table for Tortosa Storms

Transforming these counts into probabilities results in the following tables:

Exposed p-value

	No Twin	Twin	Total
Spring	0.212121	0.307692	0.247619
Fall	0.431818	0.384615	0.414286
Winter	0.356061	0.307692	0.338095
Total	1	1	1

Table 12.a: Exposed p- value

Outcome p-value

	No Twin	Twin	Total
Spring	0.538462	0.461539	1
Fall	0.655172	0.344828	1
Winter	0.661972	0.338028	1
Total	0.628571	0.371429	1

Table 12.b: Outcome p-value

Table 12: P-values of Tortosa Data

The exposed p-value in **Table 12.a** is based on the distribution of the response variable amongst all the seasons. The value in each cell can be calculated by dividing the count by the total number of storms of that particular form (Twin or No Twin). Each cell value indicates the percentage of storms that occurred given the season. The outcome p-value shown in **Table 12.b** shows the opposite. It displays the percentage of storms allocated in each season given the type of storm. These values are found by divided each cell count by the total number of storm that occurred within the given season.

The information in the above tables is used to calculate the odds ratio. This is done in order to test the strength of association of the response variable within each categorical season. Specifically, the *Fisher exact test* is applied. The odds ratio is presented along with its 95% confidence interval.

		95% Confidence Interval	
	Estimate	Lower	Upper
Spring	1	NA	NA
Fall	0.616222	0.2873078	1.317539
Winter	0.59829	0.2681752	1.32511

Table 13: Odds Ratio (Spring-base)

It should be noted in this table that the odds ratio is just an estimate, as it is labeled; however, the 95% confidence level is exact.

As mentioned before, the purpose of this test is to check the association between the seasons. In order to do this, "R" has applied three tests to the data represented in the following table:

	Midp.exact	**Fisher.exact**	Chi.square
Spring	NA	NA	NA
Fall	0.1793542	0.2089405	0.1719163
Winter	0.1729935	0.1925843	0.1653845

Table 14: P-value comparison (Spring-base)

Midp.exact and the *Fisher.exact* test both give values based on the maximum likelihood estimator. The *Chi.square* test gives an estimated value because of the limited sample size (Chi-square is better suited for large sample sizes). From all these tests, however, the values are the same in both fall and winter (to the second decimal). This is an indication that there exists no difference between the seasons in terms of frequency of twins.

3.2 Blanes

The next area to analyze is Blanes. A similar procedure is conducted in this area. The following tables show the descriptive statistics collected from the buoy:

Storm Wave Height			
Statistical Description	All	Twin	No Twin
No. of storms	69	17	47
Median	240	239	241
Standard Deviation	62	70	60
Average	251	259	248
Mode	258	239	258
Range	320	249	320
Skew	2	1	2

* : occurrence of a twin is indeterminate

Storm Duration			
Statistical Description	All	Twin	No Twin
No. of storms	69	17	47
Median	16	18	16
Standard Deviation	19	13	21
Average	23	22	24
Mode	12	12	12
Range	110	42	110
Skew	2	1	2

* : occurrence of a twin is indeterminate

Table 15: Statistics for Blanes Storms

It should be noted that, due to poorer quality of the data, some measurements showed an obvious occurrence of a storm, but whether it was followed by another could not be determined. To avoid any assumptions, those storms are label as NA and are disregarded from farther analysis.

The height variation between a storm with and without a twin is more prevalent here than in Cap Tortosa. The median is quite similar, but when examining the average, it seems that twin storms produce a higher wave height than single storm events. In terms of storm duration, judging from the two areas, it does seem that, on average, twin storms tend to be a bit shorter than non-twins (by two hours in both cases). Applying the logistic regression model as was done in the Cap Tortosa, with date being sinusoidal, displays a relation on the cosine term of the date to have a p-value of 0.065, which falls into the 0.10 significance range. Table X shows the result:

	Degree of Freedom	Deviance	Residual Degree of Freedom	Residual Deviance	P=(>|Chi|)
NULL	63	74.094			
BlanesSin	1	0.3806	62	73.713	0.53728
BlanesCos	1	3.4124	61	70.301	0.06471
BlanesHeight	1	0.7671	60	69.534	0.3811
BlanesDuration	1	0.0006	59	69.533	0.97976

Table 16: Blanes ANOVA and Chi-square with Dates

This shows a dependence on the date that was not seen in Cap Tortosa data. As before, the storms are categorized into their season of occurrence. The seasonal statistics are:

Statistical Description	All	Spring	Summer	Fall	Winter
No. of storms	69	12*	2	27*	28*
No. of Twins	17	1	0	8	8
No. of Not Twins	47	9	2	17	19
Median	240	250	185	258	226
Standard Deviation	62	33	18	71	60
Average	251	250	185	268	239
Mode	258	-	-	213	195
Range	320	121	25	257	316
Skew	2	0	-	1	3

Table 17.a: Statistics for Blanes (Seasonal) Storm Wave Height

* : occurrence of a twin is indeterminate

Seasonal Storm Duration					
Statistical Description	All	Spring	Summer	Fall	Winter
No. of storms	69	12*	2	27*	28*
No. of Twins	17	1	0	8	8
No. of Not Twins	47	9	2	17	19
Median	16	14	14	18	16
Standard Deviation	19	13	6	23	17
Average	23	19	14	27	22
Mode	12	12	-	12	8
Range	110	42	9	110	78
Skew	2	1	-	2	2

Table 17.b: Statistics for Blanes (Seasonal) Storm Wave Duration

* : occurrence of a twin is indeterminate

Table 17: Statistics for Blanes

As before, summer provides no informative contribution to the analysis. However, in this location, the same can be said for spring. This indicates that the seasonal variable will probably not provide much influence on the response, which is shown true when the model is tested with the ANOVA and Chi-square test. The tables containing the specific values as well as the mosaic plot can be found in Appendix C.

3.3 Roses

As with the data collected at Blanes, the Roses data present no dependence on any factors, neither date nor season. The statistics are represented here, but if farther data regarding this area is of interest, the results of the logistic regression, ANOVA, and Chi-square tests for both dates and seasons can be found in the Appendix D.

Storm Wave Height			
Statistical Description	All	Twin	No Twin
No. of storms	91	20	71
Median	248	248	248
Standard Deviation	101	64	108
Average	279	259	285
Mode	205	200	268
Range	431	219	431
Skew	1	1	1

Table 18.a: Statistics for Roses Storm Wave Height

Storm Duration			
Statistical Description	All	Twin	No Twin
No. of storms	91	20	71
Median	20	16	21
Standard Deviation	20	20	20
Average	26	23	27
Mode	6	18	6
Range	96	72	96
Skew	1	1	1

Table 18.b: Statistics for Roses Storm Duration

Table 18.a and Table 18.b: Statistics for Roses Storms

Seasonal Storm Wave Height					
Statistical Description	All	Spring	Summer	Fall	Winter
No. of storms	91	19	1	36	35
No. of Twins	20	5	0	10	5
No. of Not Twins	71	14	1	26	30
Median	248	239	214	236	276
Standard Deviation	101	63	-	120	96
Average	279	250	214	289	287
Mode	205	200	-	205	206
Range	431	259	0	405	405
Skew	2	1	-	2	1

Table 19.a: Statistics for Roses (Seasonal) Storm Wave Height

Seasonal Storm Duration					
Statistical Description	All	Spring	Summer	Fall	Winter
No. of storms	91	19	1	36	35
No. of Twins	20	5	0	10	5
No. of Not Twins	71	14	1	26	30
Median	20	25	11	18	23
Standard Deviation	20	14	-	18	23
Average	26	24	11	23	31
Mode	6	30	-	6	6
Range	96	51	0	93	73
Skew	1	0	-	2	1

Table 19.b: Statistics for Roses (Seasonal) Storm Wave Duration

Table 19: Statistics for Rose

3.4 Llobregat

The data collected from the buoy at Llobregat contained both higher quality and more consistent readings than did the buoys at Blanes and Roses. It is hopeful that if a trend of twin storm occurrence could exist, it would be represented by the data acquired in Llobregat. The descriptive statistics are as follows:

Storm Wave Height			
Statistical Description	All	Twin	No Twin
No. of storms	172	38	134
Median	221	221	221
Standard Deviation	53	54	52
Average	235	234	236
Mode	189	192	203
Range	297	223	297
Skew	2	2	1

Table 20.a: Statistics for Llobregat Storm Wave Height

Storm Duration			
Statistical Description	All	Twin	No Twin
No. of storms	172	38	134
Median	14	16	13
Standard Deviation	16	22	14
Average	20	24	19
Mode	12	12	12
Range	93	93	72
Skew	2	2	2

Table 20.b: Statistics for Llobregat Storm Duration

Table 20: Statistics for Llobregat Storms

The storm wave heights are remarkably similar for the three categories of information, for example, the average only differing by one centimeter and the median being the exact for all three categories. The duration of the storms seems to follow no pattern; the average duration being longer for a twin than for a non-twin, which is contradictory to the other areas' data.

Inputting the data into "R" and running an ANOVA and Chi-square on the linear regression model yields the following results:

	Degree of freedom	Deviance	Residual Degree of freedom	Residual Deviance	P(>|Chi|)
NULL			171	181.66	
LlobregatSin	1	0.0250	170	181.63	0.87436
LlobregatCos	1	7.7605	169	173.88	0.00534
LlobregatHeight	1	0.0033	168	173.87	0.95413
LlobregatDuration	1	2.7374	167	171.13	0.09802

Table 21: Anova and Chi-square test for Llobregat Data

Here, it is seen that the cosine term of the date does, in fact, contain a p-value less than the significant level. Looking at the ellipse of this provides farther insight into the correlation:

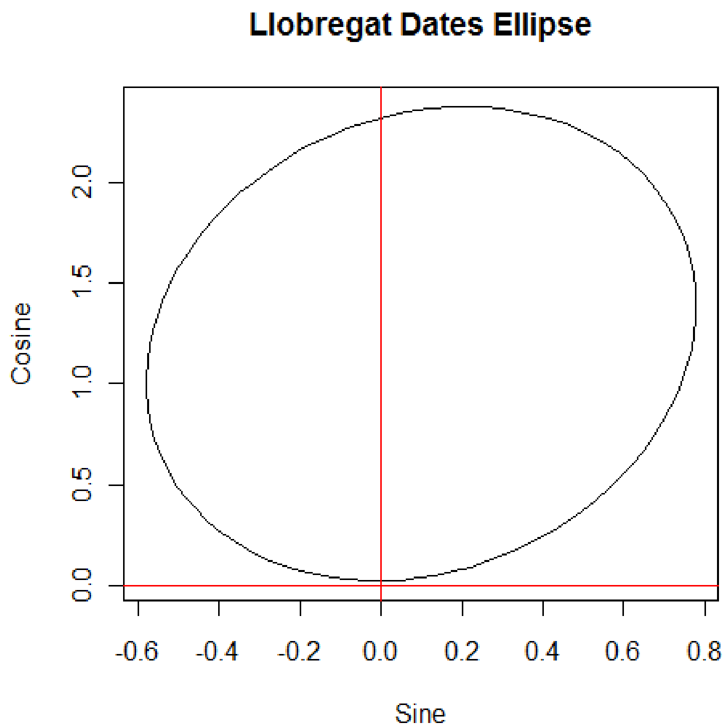

Figure 7: Llobregat Sine and Cosine Ellipse

The plot shows that the cosine term is barely out of the framework, which confirms that it is a significant term.

In addition to the significance of the cosine term, the 'duration' term misses the significance level of .05, but it does fit the 0.10 significance level. This only confirms the conclusion made when evaluating the descriptive statistics; the duration of the storm does vary, slightly, among the twin storms and the single storms.

Performing the same analysis but now applying the seasons' categorical variable to the model results in the following:

	Degree of Freedom	Deviance	Residual Degree of Freedom	Residual Deviance	P=(>|Chi|)
NULL			171	181.66	
Season	3	7.414	168	174.25	0.05981
Height	1	0.0683	167	1734.18	0.79378
Duration	1	1.6667	166	172.51	0.1967
Season: Height	3	5.4856	163	167.03	0.1395
Season:Duration	3	6.2185	160	160.81	0.10145

Table 22: ANOVA and Chi-square for Llobregat with seasons

Here, the season variable does not fit the 95% confidence interval, but it does fit the 90% confidence interval. Therefore, seasons can be said to have some influence on the occurrence of a twin. The mosaic plot is displayed below. Once again, since summer displays no twins, it is not demonstrated in the plot.

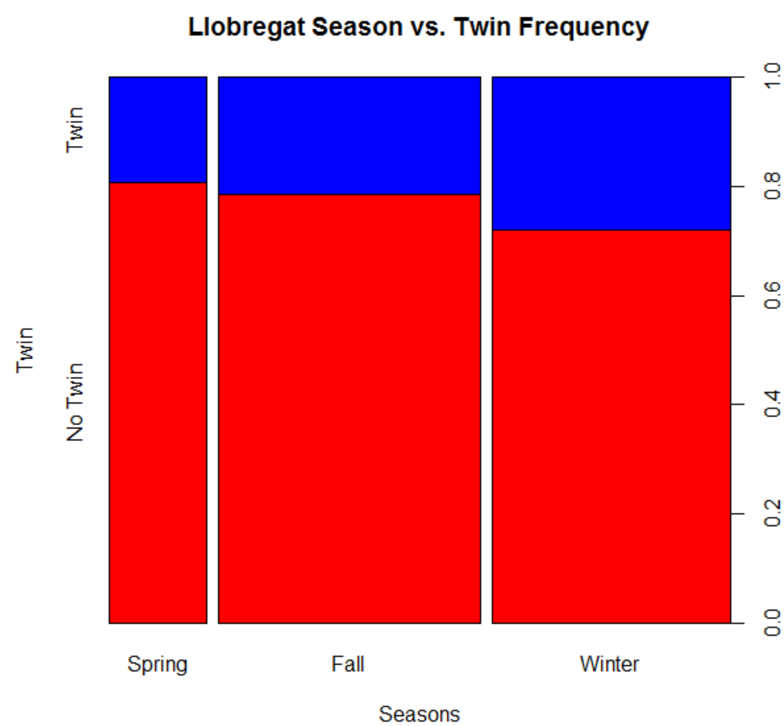

Figure 8: Llobregat Seasons Mosaic

This plot, similarly to the plot for the Cap Tortosa data, shows little variation between the seasons and the occurrence of a twin. Once again, for closer examination, the odds ratio will be studied.

Season	No Twin	Twin	Total
Spring	21	5	26
Fall	55	15	70
Winter	46	18	64
Total	122	38	160

Table 23: Contingency Table for Llobregat Storms

The p-values associated with this table are as follows:

Exposed p-value

	No Twin	Twin	Total
Spring	0.1721311	0.1315789	0.1625
Fall	0.4508197	0.3947368	0.4375
Winter	0.3770492	0.4736842	0.4000
Total	1	1	1

Table24.a: P-value (Exposed)

Outcome p-value

	No Twin	Twin	Total
Spring	0.8076923	0.1923077	1
Fall	0.7857143	0.2142857	1
Winter	0.7187500	0.2812500	1
Total	0.7625000	0.2375000	1

Table24.b: P-value (Outcome)

Table 24: Llobregat p-values

The exposed and outcome p-values have the same definitions as in the Cap Tortosa evaluation. The odds ratios of each season with respect to spring as well as the 95% confidence intervals are:

		95% Confidence Interval	
	Estimate	Lower	Upper
Spring	1	NA	NA
Fall	1.143892	0.3381990	4.534523
Winter	1.634893	0.4939341	6.402573

Table 25: Odds Ratio (Spring-base)

These results differ from the Cap Tortosa results mainly in the fact that in both fall and winter twin storms are more likely than in the spring. In Cap Tortosa, the season containing the most frequent twins was spring, resulting in odds ratio estimates less than one. Also, the confidence intervals are significantly higher in this data set, with a range of approximately four for fall and six for winter, whereas the range for Cap Tortosa was just over one for both fall and winter.

The p-values resulting from the different test are as follows:

	Midp.exact	**Fisher.exact**	Chi.square
Spring	NA	NA	NA
Fall	0.8385567	1.0000000	0.8137166
Winter	0.4003493	0.4363778	0.3805904

Table 26: P-value Comparison (Spring-base)

This also produces interesting outcomes when compared to the p-values for Cap Tortosa, where the seasonal values for each test are essentially the same. Here, the p-value of winter is approximately half of the p-value for fall.

3.5 Sensitivity Analysis

Since there seems to be no similarity in parameters based on different locations, the next step was to test how sensitive the data is to change in storm criteria. Therefore, the storm wave height threshold was changed to 200 cm for the information gathered at Cap Tortosa and Llobregat. This was not done for Blanes or Roses because the number of storms for both locations was too low to ensure accurate results when inputted into the model. Also, based on the ambiguity of the time measurements in for each decade in different locations, a change in duration criteria was not preformed. Consequently, this sensitivity analysis only focuses the sensitivity to a change in threshold storm wave height.

In Cap Tortosa, the number of storms and the probability of occurrence for a threshold of 200 cm compared to a threshold of 160 cm is represented by the following tables:

Tortosa: Threshold= **160 cm**			Tortosa: Threshold= **200 cm**		
Twin	No Twin	Total	Twin	No Twin	Total
78	138	216	15	72	87
0.36111111	0.63888889	1	0.172414	0.82756	1

Table 27: Tortosa Threshold Comparison

It can be seen that the probability of a twin drops significantly when the threshold is changed, namely, almost by half. This implies that the threshold value will influence the occurrence of a twin. Applying the data with the new threshold into the model shows dependence with a significance level of 0.064 on the cosine date term, similar as that experienced in the Blanes and the Llobregat data at a threshold of 160 cm. No dependence is shown with the model is run with the categorical seasons variable. The tables can be seen in Appendix E.

Next, the same is done for the Llobregat data. The comparison between the two thresholds is:

Llobregat: Threshold= **160 cm**			Llobregat: Threshold= **200 cm**		
Twin	No Twin	Total	Twin	No Twin	Total
38	122	160	8	63	71
0.2375	0.7625	1	0.11267606	0.88732394	1

Table 28: Llobregat Threshold Comparison

Similar to the Cap Tortosa data, the probability of a twin drops significantly in Llobregat. However, the most interesting aspect is that the probability drops almost the same amount in both locations (~ 47%). This implies that not only is there dependence on the storm wave height threshold value, but this can somehow be quantified. Inserting the test into the logistic regression model and performing the tests of independence show that the only variable that influences the twin event is the interaction of the 'Seasons' variable and the 'Duration' variable, with a significance level of 0.014. These results can also be found in Appendix E.

Appendix E also contains similar comparisons for both locations of the storms in each season. The results follow a similar pattern to the comparisons shown in **Tables 27** and **28** and will not be discussed further.

3.6 Cost and Risk Results

The twin probabilities resulting from the above analysis for each area and season are:

Probability	
Tortosa	**Llobregat**
0.361111111	0.22093023

Table 29.a: Twin Probability

Probability		
Tortosa	Spring	0.461538
	Fall	0.344828
	Winter	0.338028
Llobregat	Spring	0.192308
	Fall	0.214286
	Winter	0.28125

Table 29.b: Season Twin Probability

Table 29: Twin Probabilities for Tortosa and Llobregat

It should be noted that these probabilities include the total number of storms (including the summer storms, which are disregarded in most other probability calculations in this thesis.

Applying the overall twin probabilities for each area into the SCR_T *vs. Probability* curve mentioned before results in the following plot:

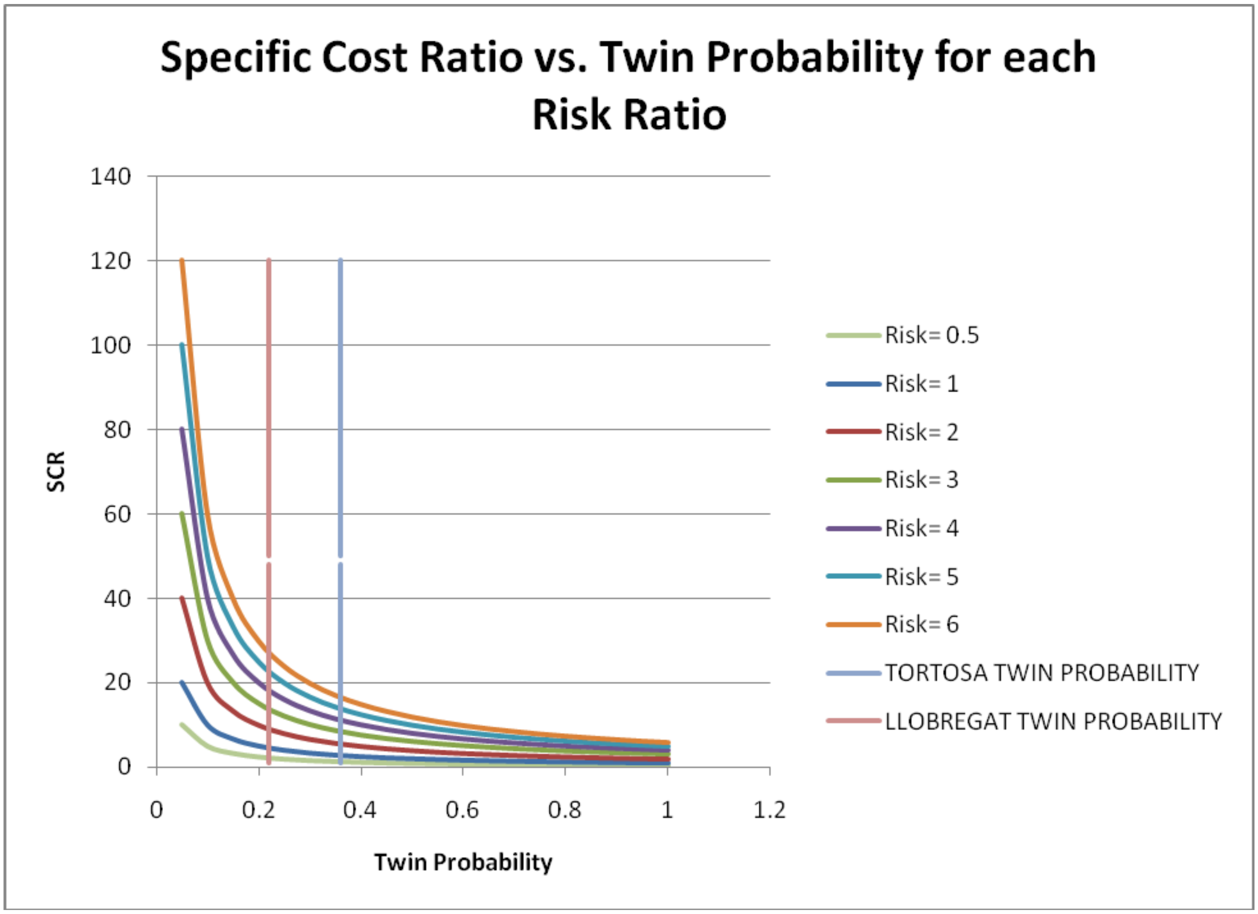

Figure 9: Cost vs. Probability for each Risk Ratio

At each probability level, the increase from one risk to the next is directly proportional to the SCR_T. Lower probabilities have larger increases in change of the cost ratios. This implies that if there is a decrease in the risk, even slight, for a small probability, acting between storms can decrease the damage significantly, whereas with a large probability, a large change in risk is needed in order for the damage cost to decrease a noticeable amount. For the probability in Tortosa (36%), each change in risk results in an approximate increase in the SCR_T of 2.9. A probability of 22% in Llobregat corresponds to an increase of 4.5.

Next, the parameters associated with these probabilities and the chosen risks will be examined more closely. For Cap Tortosa, the table below shows the probability values for all twins as well as the probability values for twins occurring in each season. From the **Table 6**, the SCR_T, β, and ρ values have been extracted for the risks that correspond to the probabilities. The tables displaying the values of β refer to a *maximum* value of β, which means that any value higher will result in a SCR_T lower than the anticipated value. The same can be said for the value of ρ in that it is a minimum value where any lower value will be out of range of the acceptable values from the cost criterion.

Tortosa				
	RR	SCR	β(max)	ρ(min)
All Twins (36%)	0.5	1.388889	0.6	0.6
	1	2.78	0.2	2.6
	1.5	4.16	0.1	5.6
Spring (46%)	0.5	1.083333	0.9	0.1
	1	2.173913	0.4	1.6
	1.5	3.26087	0.2	3.4
	2	4.347826	0.1	6
	2.5	5.434783	0.1	9.8
Fall (34%)	0.5	1.45	0.6	0.6
	1	2.941176	0.2	2.8
	1.5	4.411765	0.1	6.2
Winter (33%)	0.5	1.479167	0.6	0.6
	1	3.030303	0.2	3
	1.5	4.545455	0.1	6.6

Table 30: Tortosa ρ and β values

The maximum risk that is applicable in this area is 2.5 correlating to a probability of 0.46 and a SCR_T *of* 5.43. Therefore, in the spring, it makes sense to act between storms if acting can reduce the damages of the second storm to 20% of the damages without acting and that the cost of not acting will produce damages almost ten times greater than the damages from the first storm.

Observing the same calculations in Llobregat:

Llobregat				
	RR	SCR	β(max)	ρ(min)
All Twins (22%)	0.5	2.272727	0.3	1.8
	1	4.545455	0.1	6.4
Spring (19%)	0.5	2.6	0.3	2.2
	1	5.263158	0.1	9
Fall (21%)	0.5	2.333333	0.3	1.8
	1	4.761905	0.1	7.2
Winter (28%)	0.5	1.777778	0.5	1
	1	3.571429	0.2	4
	1.5	5.357143	0.1	9.4

Table 31: Llobregat ρ and β values

Due to the much smaller probabilities that are present in Llobregat, the maximum risks are also smaller. The most that acting between storms can lessen the risk is 1.5 times, which results from a probability of 0.28, a β value of 0.1, and a ρ of almost ten.

Conclusions

Unfortunately, the only certain conclusion that can be drawn from this analysis is that the data does not seem to present a pattern for determining twin storm events. However, similarities do exist in the results; this section will exploit these similarities in hopes of coming to some conclusions about the predictability of twin storms. Also, since the evaluation computed at Cap Tortosa and Llobregat contain the highest level of confidence in the accuracy of the measurements, only they will be evaluated. Since they both displayed a dependence on the 'Seasons' variable, the relative risk for each season in each location will be compared amongst all the others in order to see if one season has higher probability of producing a twin storm. Next, the probabilities will be implemented into the cost and risk analysis. Different adaptations of the analysis will be discussed as well as the short-comings that arise when basing decisions on the values obtained through the risk analysis.

4.1 Conclusions of the Probability Analysis

Unfortunately, there seems to be no consistency in the results regarding the variable dependence of a twin storm. However, conclusions will be made based on the similarities between the Cap Tortosa and Llobregat data, which can be validated by the fact that these areas have the most precise and consistent data readings. Therefore, the first viable conclusion is that twin storms in the areas of Blanes and Roses would also display certain behavioral characteristics if the data was more dependable. Since this is not the case, they are eliminated from any farther discussion.

Examining the results of the Cap Tortosa and Llobregat analysis leads to some important implications; the first being that storm intensity defined by wave height does not affect whether or not a twin occurs. However, a change in threshold will affect both the probability of the twin event as well as the variables on which it depends, which also varies between the two locations. Therefore, it can be said that once a certain storm intensity criterion is defined, the wave height will have no effect on the occurrence of a twin. However, the occurrence of a twin is sensitive to the criterion level chosen. The next variable, duration of the first storm, does present an influence in Llobregat, but not in Tortosa. A similar contradiction exists regarding the dependence on date. A periodic relationship has no affect in Cap Tortosa but does have a significant component in Llobregat, namely, the cosine term. The only conclusion that can be drawn from this is that a twin event could have some dependence on the date. However, since only the cosine term is significant, quantifying the relationship is challenging and given that this did not occur in Cap Tortosa, it will not be considered further. The same can be said of the 'Duration' variable in Llobregat. Because its p-value is within the 90% confidence range, it can be said that storm duration plays a role in twin storm episodes, but the lack of support from the Cap Tortosa data will dismiss its analysis from this thesis.

This implies that the only remaining variable affecting twin occurrences is the season. This has implications on the chosen probability model as well, since the 'season' variable is a categorical variable. The log-linear model now becomes the appropriate model to apply to the data.

Now that it is known that the 'Seasons' variable can be considered the only consistently-significant predictor, the question becomes which season displays a higher probability. Summer can quickly be eliminated as a possibility, since no twins were recorded in that season.

	Cap Tortosa		Llobregat	
Season	No Twin	Twin	No Twin	Twin
Spring	28	24	21	5
Fall	57	30	55	15
Winter	47	24	46	18
Total	132	78	122	38

Table 32: Contingency Tables of the Storms in Both Locations

Examining the seasonal contingency tables again shows that the 'No Twin' category is quite consistent amongst the seasons in both locations. However, twin storm events differ between locations greatly. Calculating the relative risk, where

$$RR_{Season} = P_{Twin} / P_{NoTwin} \tag{10}$$

of a twin for each season between the areas results in:

Season	Cap Tortosa	Llobregat
Spring	0.857143	0.238095
Fall	0.526316	0.272727
Winter	0.510638	0.391304
All Twins	0.590909	0.311475

Table 33: Relative Risk

In each season, the relative risk of experiencing a twin is higher in Cap Tortosa than in Llobregat. The most notable difference is found in spring due to the large relative risk of a twin in Cap Tortosa, whereas in Llobregat, the opposite occurs. The smallest relative risk of a twin arises in spring in Llobregat. Also, the difference in relative risk between fall and winter in Cap Tortosa is significantly smaller than the difference in the two seasons in Llobregat. Overall, analyzing the relative risks of a twin occurring in each season in both locations does not offer any suggestion that the presence of a twin storm event can be more or less likely during a particular season. Examining the odds ratios for the two areas will indicate the same result.

Performing the ANOVA and Chi-square tests on the *GLM* suggests that the occurrence of a twin is dependent on the season variable. After eliminating summer from the analysis and comparing the statistics for each location, no conclusion can be drawn on the actual influence that the season may have. The Cap Tortosa data emphasizes that the occurrence of a twin in spring is more probable than in fall or winter, whereas the Llobregat data advocates differently. This implies that the dependence may rely on some other unexplored factor, such as wind and wave direction. Wave direction may be more prominent in specific seasons and can influence a twin event.

4.2 Conclusions regarding Cost and Risk Analysis

The application of this cost and risk analysis should ultimately lead to a procedure in determining when acting between twin storm events is practical. Suppose risk analysts were to take the above study areas and determine if (or when) a twin storm occurrence is a valid hazard. These two study areas, for instance, have a very high value to the Catalonia due to the presence of the two ports discussed before (Port of Barcelona and Port of Tarragona, respectively represented by the Llobregat and Cap Tortosa buoys). A possible situation would occur if the breakwaters experience damage during one storm and cannot adequately protect the harbor against advances of the second storm. Since the port traffic is high in both, damage to the breakwaters has to be kept at a minimum in order to keep traffic flowing as usual.

The acceptable risk, *R*, is assigned a certain value for this area. Also, the damage of the first storm, C_0, is assumed to be known based on prediction patterns related to the seasons, wave direction, etc… Furthermore, the cost of the second storm, C_1, can be estimated. For the sake of this example, assume the damages from the second storm are twice as bad as the damages from the first storm ($\rho = 2$). Coastal engineers think of a method for minimizing the damages of the second storm by 50 %, ($\beta = 0.5$). The values of ρ and β are then applied to find the SCR_T **,** which, in this case, is equal to 1.5. If the value of *R* is set, then a probability can be determined using the risk equation. If the acceptable risk is a range of values, then an array of probabilities will result. It then becomes a matter of comparing the calculated probabilities with those listed in **Table 32** and determining if, in fact, a twin storm event is a possible hazard, which can then lead to a decision to act or not.

If it is the desire of the risk analyst to determine when action is beneficial based on a sufficiently large damage reduction between acting and not acting, applying the same values of ρ, C_0, and C_1 and the probability values from **Table 32**, the value of β could be extracted from the risk equation. Then, a decision based on the feasibility of this damage reduction could be determined.

One unfortunate aspect of this analysis is that it is insensitive to the severity of possible damage, meaning, it is has almost no dependence on ρ. In the situation regarding damage to breakwaters protecting Catalan harbors, whether the damages are presumed to be significant or not does not noticeably change the risk. An example would be if a twin event were to hit the Port of Tarragona in spring with the second storm having a capacity to be 20 times more destructive than the first storm or if it had the capacity to be 10 times more destructive the first. It should be kept in mind that this destructive capacity is not solely based on the hazard of the second storm, which may be less than the hazard of the first storm. It is also dependent on the increased vulnerability of the entire port system due to the exposure induced by the first storm. In the first situation (ρ = 20), assuming a β of 0.5 would result in a risk of 0.87. Applying the second value (ρ = 10) at the same β would result in a risk of 0.84. However, when taking damages in terms of millions of Euros, a difference in ρ being 20 and ρ being 10 is quite significant. This demonstrates that essentially, the risk associated with twin storms is dependent on β. Therefore, the risk could display a result that would render one to think that acting would not be worthwhile when this is evidentially not the case.

Another reason why this risk evaluation may not be the most effective tool for determining when to mitigate damage caused by the second storm stems from the concept of indirect consequences. For example, during the fall seasons in Llobregat, the probability of a twin is 0.21 with a corresponding risk of 1. However, when examining ρ, it shows that the damage from the second storm could be seven times the damage of the first storm. Now, taking into account indirect consequences, this value may be unacceptable. Also, this risk value of 1 demands a β as low as 0.01. Dropping the damage costs of the second storm by 90% may be an unreasonable request. This, as well as the situation above, goes to show that the decision to implement preemptive measures cannot be based solely on the analysis presented here.

Suggestions for Further Analysis

The conclusions of this thesis only go so far as to suggest that twin storm events are in some way dependent on the season. This section will mention the aspects of the analysis that could be explored more thoroughly and make some suggestions regarding the advancement of the investigation for a better prediction system. As mentioned before, considering other factors such as wave or wind direction may shed some light on how the season in which an event occurs can influence the occurrence of a twin storm. Wave period could have some influence as well. Also, in the data acquired from the Llobregat buoy, a relation between the cosine term of the date function proved significant. The 'duration' term in the *GLM* displayed importance as well. Both components could render further investigation.

When analyzing the details of the first storm, the storm intensity measurement is limited only to storm wave height, whereas other storm characteristics such as storm surge or wind and wave set-up can very much influence the destructiveness of storms. Therefore, the risk analysis could benefit from including these factors in the description of the first storm. In terms of the second storm, the previous conclusions would benefit greatly if the intensity of the second storm is more closely analyzed. In this report, details of the second storm were not considered outside of determining whether the wave height and duration fit the criteria of a twin storm. Investigating the details of the second storm can shed some light on the expectations associated with its destructive characteristics.

Another aspect to consider is the effects of climate change and if this could possibly influence the occurrence of twin storms in the Mediterranean Sea. Scientists have hypothesized that storms can possibly increase in frequency and intensify during the coming years as a side-effect of the global warming. However, in reference to the conclusions achieved in this analysis with regards to the *hazard* of a twin storm, increase in storm frequency and intensity should not have an influence because neither of the two should induce any change in the pattern of twin events due to the fact that the pattern does not seem to be related to storm wave height or duration. However, it can have a very large effect on the consequences related to twin storms which constitutes it as an important aspect to consider in the risk analysis.

Though the cost and risk analysis can provide some information on how to determine which probabilities would suggest that acting between twins could be cost efficient, there are still many questions that need to be considered. For instance: *How likely is it to have a twin storm event where the damage of the second storm is significantly higher than the first storm?* Another arising question becomes *are there methods and practices that can be instilled so that a high required cost reduction between acting and not acting is feasible?* These possibilities can severely limit the application of the analysis. Detailed answers to these questions should be investigated on a site-specific basis, since the answers are highly dependent on the location of interest.

In addition, the only examples provided here are in regards to the Ports of Barcelona and Tarragona. There exist many other consequences that can result due to wave damage, such as beach width loss, damage to coastal structures such as groins and jetties, or damage to structures located on or near the beach front. Performing evaluations based on these areas as well can provide some insight in the usefulness of this study.

Concluding Remarks

In conclusion, the analysis of a twin storm event renders farther investigation. This study has shown that the probability of occurrence of a twin as well as the damage that can result both provide strong evidence to generate concern on the issue. If an accurate prediction system can be developed and used in conjunction with cost- effective and efficient mitigating measures, the reduction in damage to the beaches and structures located along the coast can be greatly minimized. With the present and growing demands for a stable coastline, this is of upmost importance to engineers and coastal zone managers alike. Taking into consideration the above-mentioned suggestions in combination with more specific, longer lasting data could increase the accuracy of the results and thus prelude more concrete conclusions. However, based on this analysis, it can be said for sure that the Catalan-Mediterranean coast does display a trend of twin storm events.

References

Agresti, Alan (2007). *An Introduction to Categorical Data Analysis.* 2nd ed. New York, Wiley

Agresti, Alan (2002). *Categorical Data Analysis.* 2nd ed. Hoboken [N.J.]: John Wiley & Sons

Bowman, D., Guillén,J., López L. and Pellegrino, V. (2009) *Planview Geometry and Morphological Characteristics of Pocket Beaches on the Catalan Coast (Spain)*, Geomorphology, Volume 108, Issues 3-4, 191-199, July 2009.

Gaur, Ajai S., Gaur, Sanjaya S. (2009). *Statistical Methods for Practice and Research*. 2nd Edition. New Delphi: Response Books.

Hastie, T. J., Pregibon, D. (1992). *Statistical Models in S*: Ch. 6: *Generalized linear models.* J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole, 1992.

Jiménez, J., Ciavola, P., Balouin, Y., Armaroli, C., Bosom, E., and Gervais, M. (2009). *Geomorphic coastal vulnerability to storms in microtidal fetch-limited environments: application to NW Mediterranean & N Adriatic Seas.* Journal of Coastal Research, Special Issue 56, 2009.

Jiménez J.A., Sánchez-Arcilla A., Valdemoro H.I., Gracia V. and Nieto F., (1997). *Processes Reshaping the Ebro Delta*, Marine Geology, 144, pp. 59- 79

Llasat, M.C., LLasat-Botija, M., Barnollas, M., Lopez, L., Altava-Ortiz, V. (2009). *An analysis of the evolution of hydrometeorological extremes in newspapers: the case of Catalonia, 1986-2006*. Natural Hazards Earth System Sciences.,9, 1201-1212, 2009.

Llasat, M.C., LLasat-Botija, M., and Lopez, L.(2009). *A press database on natural risks and its application in the study of floods in Northeastern Spain*. Natural Hazards Earth System Sciences.,9, 2049-2061, 2009.

McCullagh, P., Nelder,J. A. *Generalized Linear Models*. 2nd ed. London: Chapman and Hall, 1989.

Mendoza, E**.,** Jiménez, J. (2008). *Coastal Vulnerability to Storm in the Catalan Coast. UPC Doctorate Dissertation.* Available at: http://www.tesisenxarxa.net/TDX-0414108-122517/

Mendoza, E. T. & Jimenez, J. A. (2008). *Regional vulnerability analysis to storm impacts in the Catalan Coast.,* Proceedings of the Institution of Civil Engineers: Maritime Engineering, (in review).

Mendoza, E.T. , Jiménez, J.A. (2008). Vulnerability Assessment to Coastal Storms at Regional Scale. Proc. 31st International Coastal Engineering Conference, ASCE, Hamburg, 4154-4166, 2008.

Port of Barcelona information available at: www.portdebarcelona.es, 2010.

Port of Tarragona information available at: www.porttarragona.es, 2009.

Puertos de Catalunya, avaliable at: www.capitanes.com , 1998 - 2009.

Sanchez-Arcilla, A., Gonzalez-Marco, D., Bolanos, R. (2008). *A Review of Wave Climate and Prediction along the Spanish-Mediterranean Coast.* Natural Hazards Earth System Sciences, 8, 1217-1228, 2008.

UNISDR, 2009: UNiSDR Terminology on Disaster Risk Reduction, available at: http://www.unisdr.org/eng/erminology/terminology-2009-eng.html.

Verhagen, H.J., d'Angremond, K. van Roode, F. (2009). *Breakwaters and Closure Dams.* VSSD, Delft, The Netherlands 2009.

Appendix A: Cap Tortosa Data

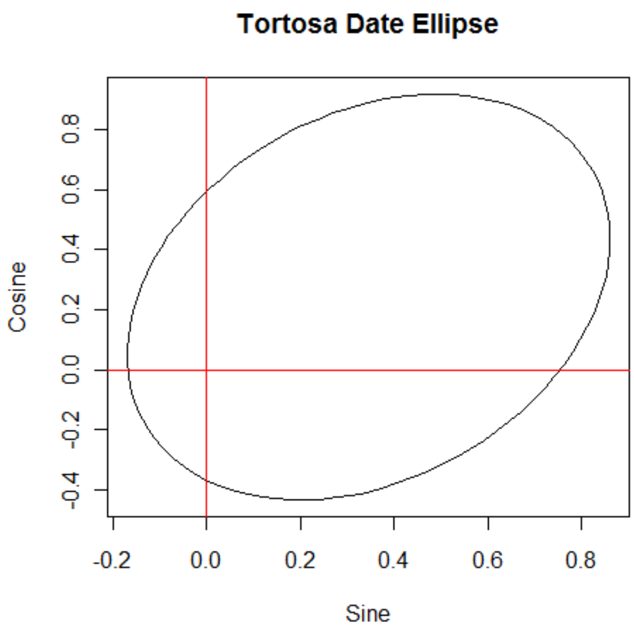

Figure 10: Cap Tortosa Date Ellipse

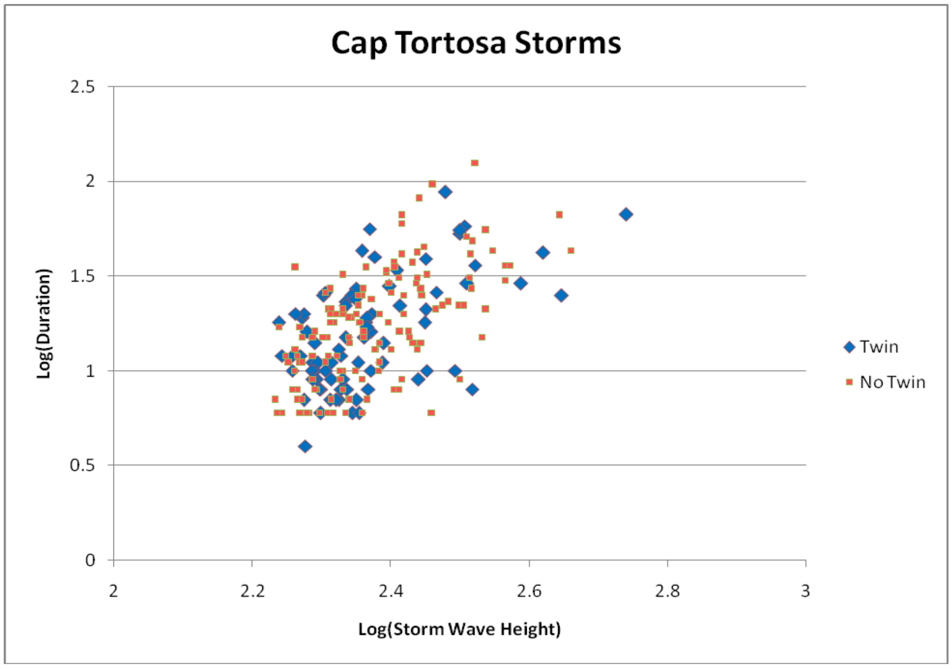

Figure 11: Storm Wave Height vs. Duration

Appendix B: Llobregat Data

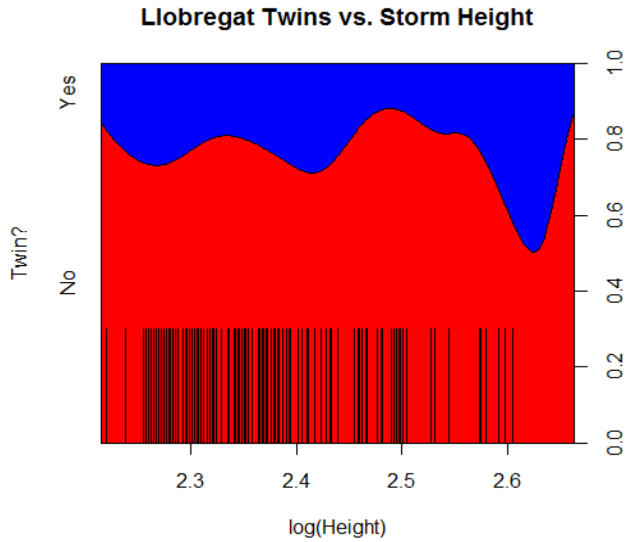

Figure 12.a: Storm Height Conditional Density

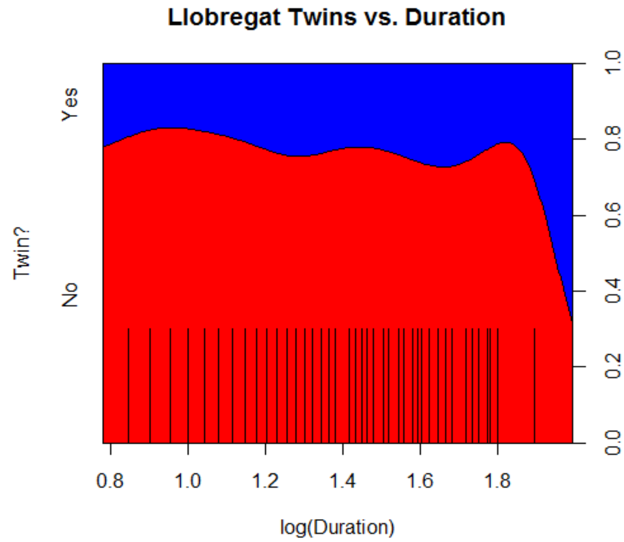

Figure 12.b: Duration Conditional Density

Figure 12: Llobregat Conditional Density Plots

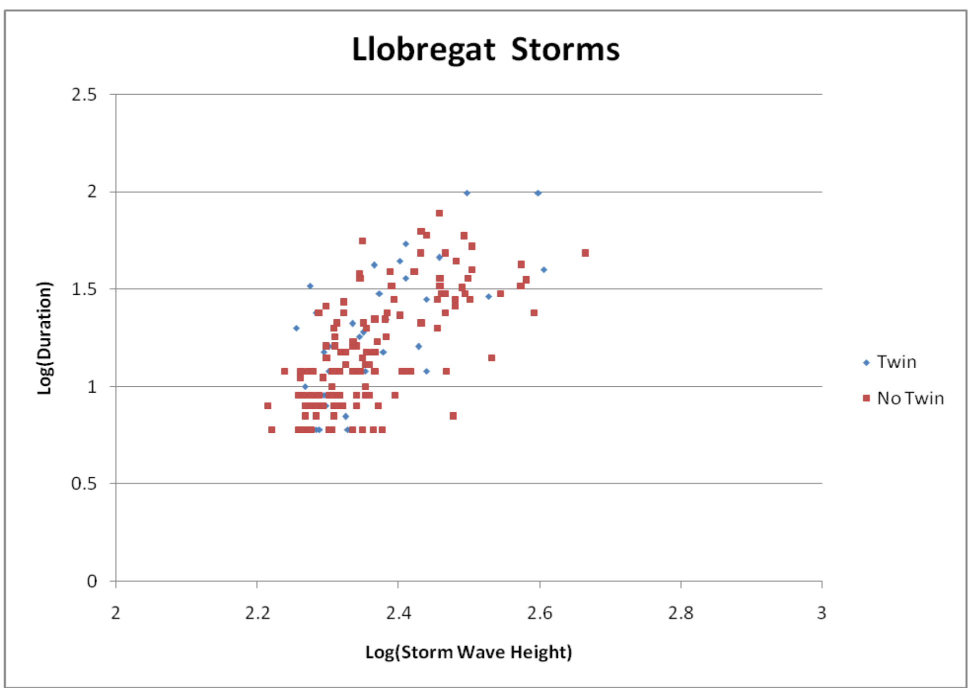

Figure 13: Storm Wave Height vs. Duration

Appendix C: Blanes Data

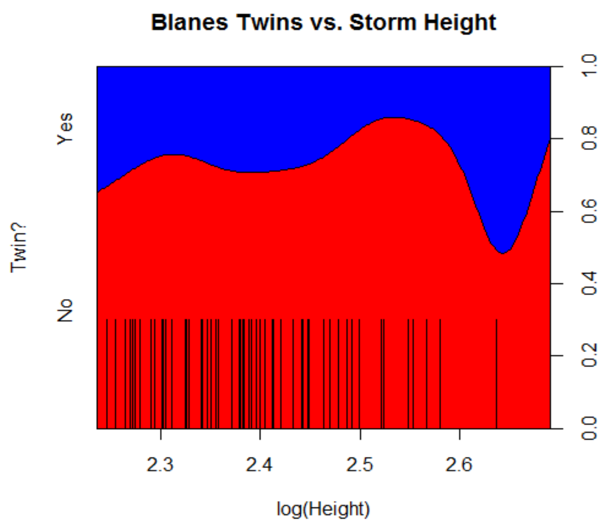

Figure 14.a: Storm Height Conditional Density

Figure 14.b: Duration Conditional Density

Figure 14: Blanes Conditional Density Plots

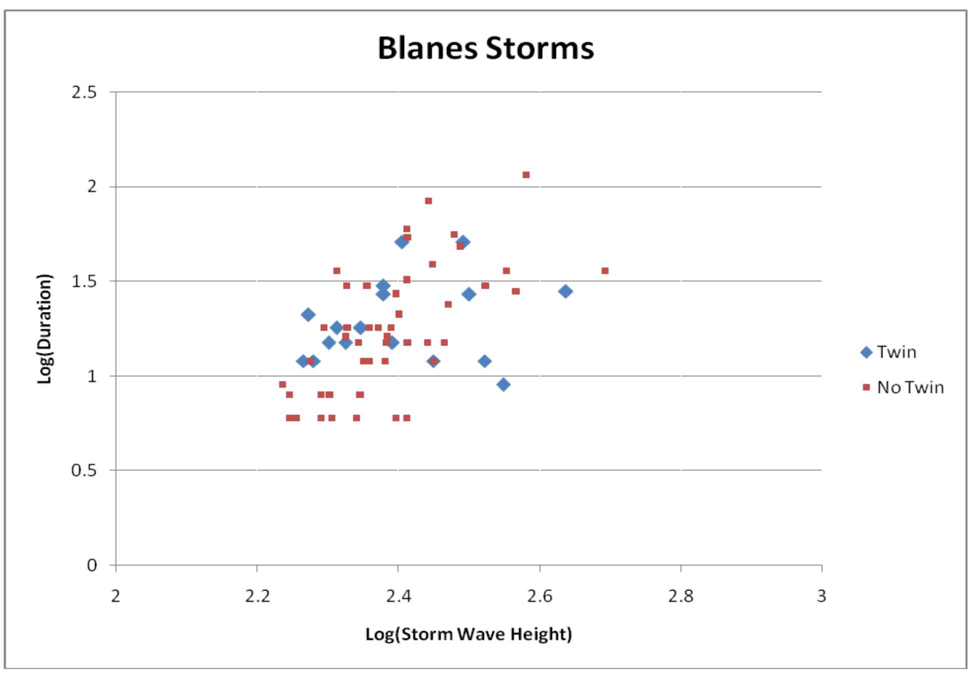

Figure 15: Storm Wave Height vs. Duration

	Degree of Freedom	Deviance	Residual Degree of Freedom	Residual Deviance	P=(>|Chi|)
NULL	63	74.094			
BlanesSeason	1	1.76322	62	72.331	0.1842
BlanesSHs	1	0.2998	61	72.031	0.584
BlanesSD	1	0.00001	60	72.031	0.9974
BlanesSeason:BlanesSHs	1	0.43942	59	71.591	0.5074
BlanesSeason:BlanesSD	1	2.13879	58	69.453	0.1436

Table 34: Blanes ANOVA and Chi-square with Seasons

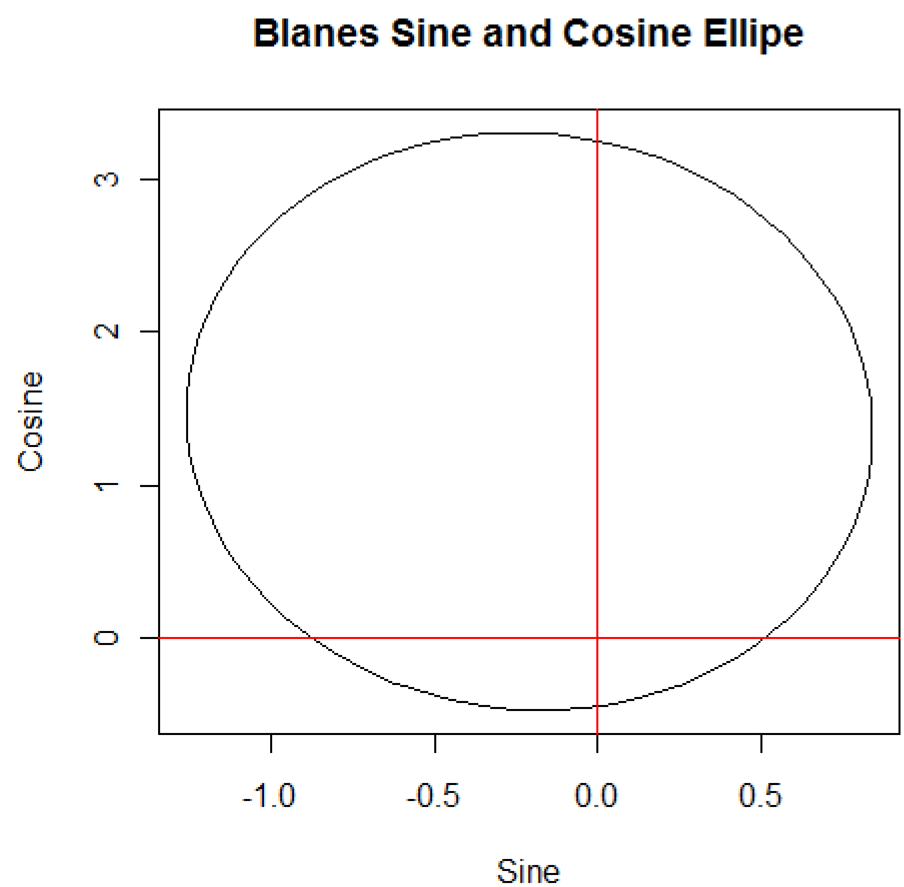

Figure 16: Blanes Ellipse

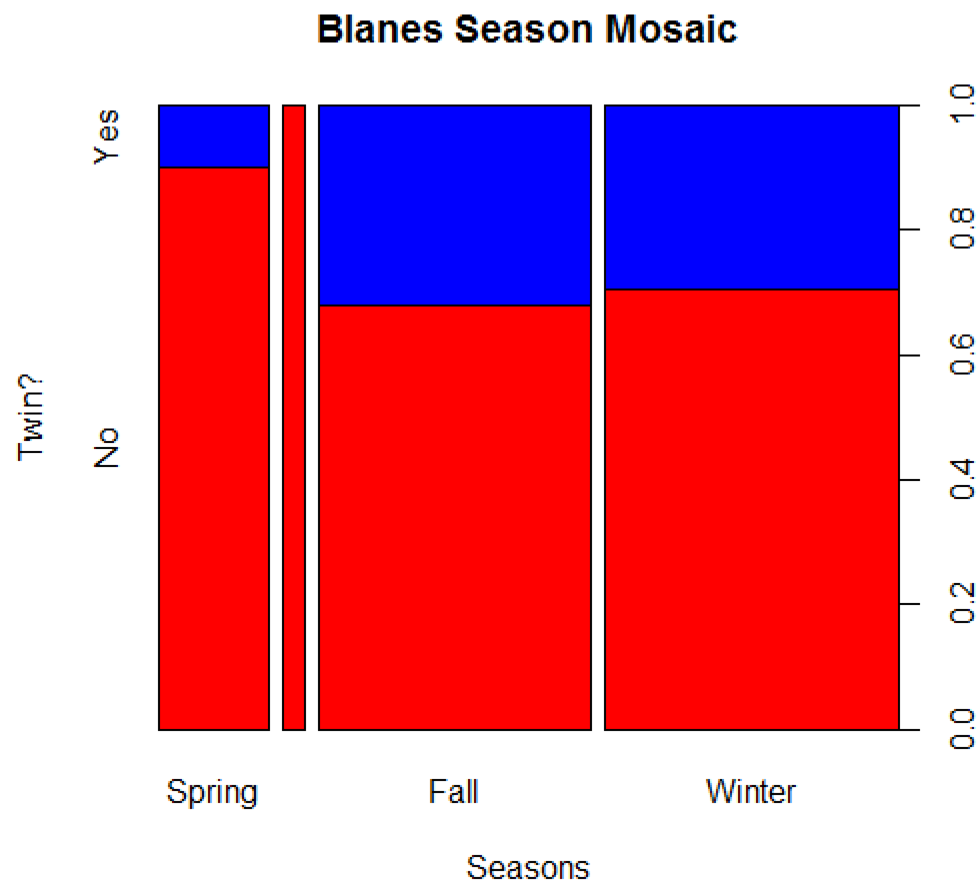

Figure 17: Blanes Seasons Mosaic

Appendix D: Roses

Seasonal Storm Wave Height					
Statistical Description	All	Spring	Summer	Fall	Winter
No. of storms	91	19	1	36	35
No. of Twins	20	5	0	10	5
No. of Not Twins	71	14	1	26	30
Median	248	239	214	236	276
Standard Deviation	101	63	-	120	96
Average	279	250	214	289	287
Mode	205	200	-	205	206
Range	431	259	0	405	405
Skew	2	1	-	2	1

Table 35.a: Statistics for Roses (Seasonal) Storm Wave Height

Seasonal Storm Duration					
Statistical Description	All	Spring	Summer	Fall	Winter
No. of storms	216	52	6	87	71
No. of Twins	78	24	0	30	24
No. of Not Twins	138	28	6	57	47
Median	20	25	11	18	23
Standard Deviation	20	14	-	18	23
Average	26	24	11	23	31
Mode	6	30	-	6	6
Range	96	51	0	93	73
Skew	1	0	-	2	1

Table 35.b: Statistics for Roses (Seasonal) Storm Wave Duration

Table 35: Roses Seasonal Descriptive Statistics

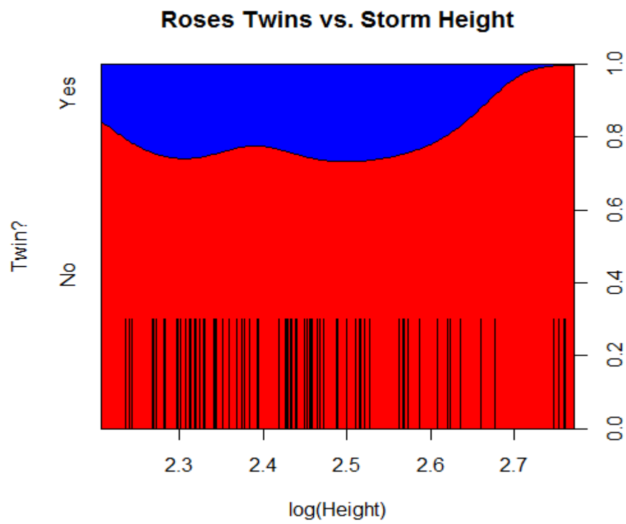

Figure 18.a: Storm Height Conditional Density

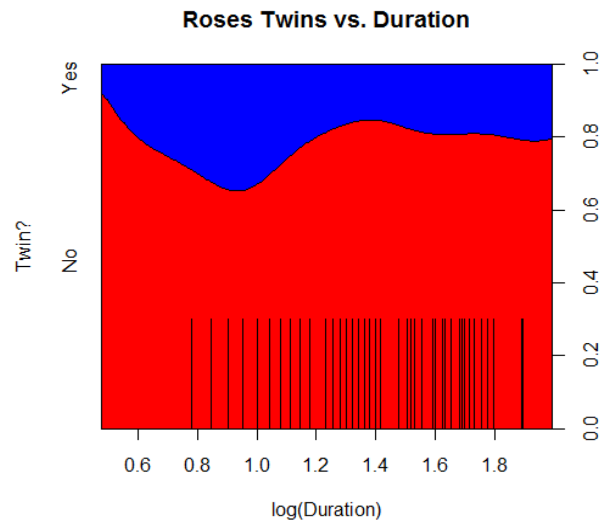

Figure 18.b: Duration Conditional Density

Figure 18: Roses Conditional Density Plots

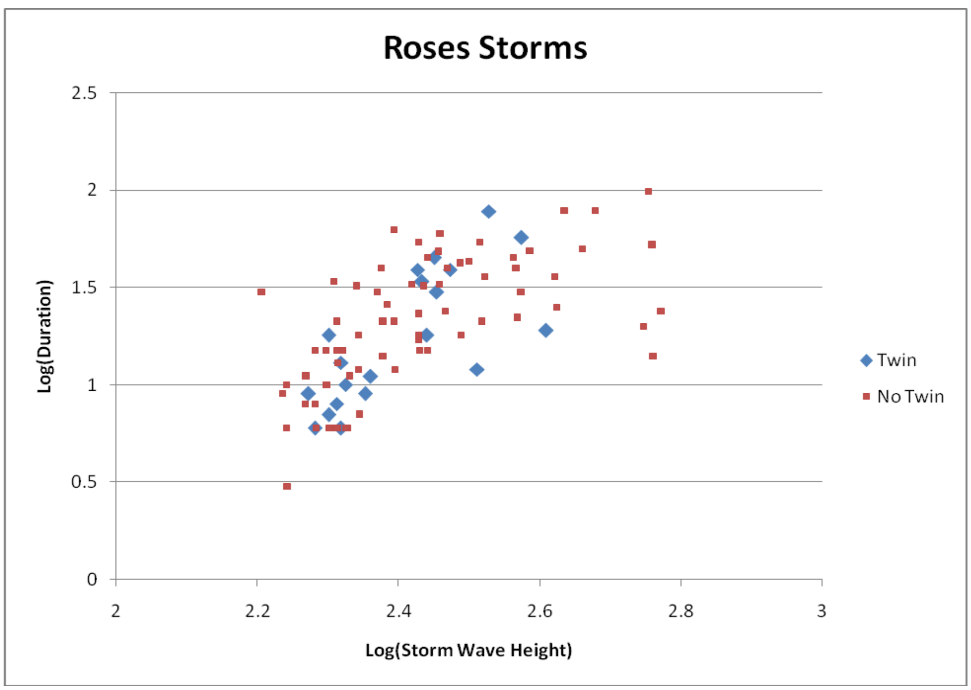

Figure 19: Storm Wave Height vs. Duration

	Degree of Freedom	Deviance	Residual Degree of Freedom	Residual Deviance	P=(>|Chi|)
NULL	90	95.847			
RosesSin	1	0.40977	89	95.437	0.5221
RosesCos	1	0.00035	88	95.436	0.9851
RosesHeight	1	0.75801	87	94.678	0.384
RosesDuration	1	0.14622	86	94.532	0.7022

Table 36: Roses ANOVA and Chi-square with Dates

	Degree of Freedom	Deviance	Residual Degree of Freedom	Residual Deviance	P=(>|Chi|)
NULL	90	95.847			
RosesSeason	3	2.69712	87	93.149	0.4407
RosesHeight	1	0.58344	86	92.566	0.445
RosesDuration	1	0.07869	85	92.487	0.7791
RosesSeason:RosesHeight	2	0.59725	83	91.89	0.7418
RosesSeason:RosesDuration	2	0.74978	81	91.14	0.6874

Table 37: Roses ANOVA and Chi-square with Seasons

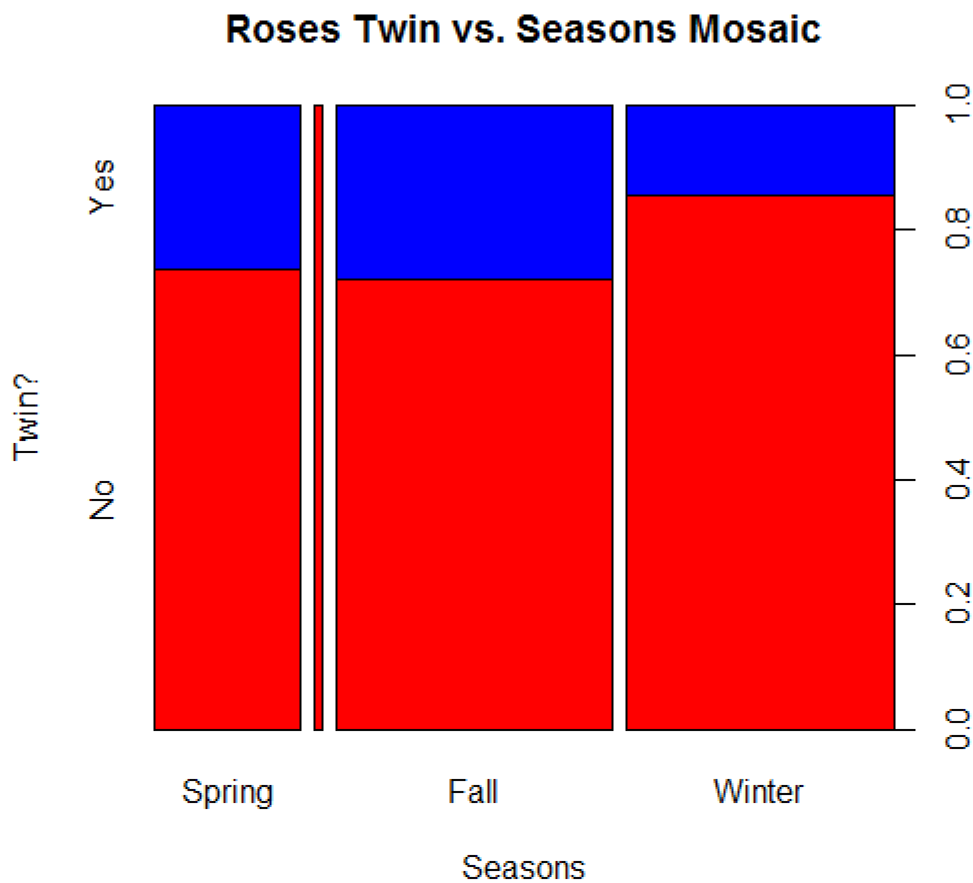

Figure 20: Roses Season Mosaic

Appendix E: Sensitivity Analysis

TORTOSA

Independence Test Results

	Degree of Freedom	Deviance	Residual Degree of Freedom	Residual Deviance	P=(>|Chi|)
NULL	86	79.987			
logHs	1	1.6977	85	78.289	0.19259
logD	1	0.3941	84	77.895	0.53016
sinDate	1	1.0148	83	76.88	0.31375
cosDate	1	3.5798	82	73.3	0.05849

Table 38: ANOVA and Chi-square with Dates (Threshold= 200 cm)

	Degree of Freedom	Deviance	Residual Degree of Freedom	Residual Deviance	P=(>|Chi|)
NULL	86	79.987			
logHs	3	4.8017	83	75.185	0.1869
logD	1	1.7143	82	73.471	0.1904
Season	1	0	81	73.471	0.9959
Season:logHs	3	2.3971	78	71.073	0.4942
Season:logD	3	4.0645	75	67.009	0.2546

Table 39: ANOVA and Chi-square for Tortosa with Seasons (Threshold= 200 cm)

Contingency and Probability Tables

Season	No Twin	Twin	Total
Spring	28	24	52
Fall	57	30	87
Winter	47	24	71
Total	132	78	210

Table 40.a: Contingency Table for Tortosa Storms (Th= 160 cm)

Season	No Twin	Twin	Total
Spring	17	2	19
Fall	28	10	38
Winter	23	3	26
Total	72	15	87

Table 40.b: Contingency Table for Tortosa Storms (Th= 200 cm)

Table 40: Contingency Tables for Tortosa

Season	No Twin	Twin	Total
Spring	0.133333	0.114286	0.247619
Fall	0.271429	0.142857	0.414286
Winter	0.22381	0.114286	0.338095
Total	0.628571	0.371429	1

Table 41.a: Probability Table for Tortosa Storms

Season	No Twin	Twin	Total
Spring	0.195402	0.022989	0.218391
Fall	0.321839	0.114943	0.436782
Winter	0.264368	0.034483	0.298851
Total	0.827586	0.172414	1

Table 41.b: Probability Table for Tortosa Storms (Th= 200 cm)

Table 41: Probability Tables for Tortosa

LLOBREGAT

Independence Tests Results

	Degree of Freedom	Deviance	Residual Degree of Freedom	Residual Deviance	P=(>|Chi|)
NULL	71	50.232			
SinDate	1	0.0832	70	50.149	0.77302
CosDate	1	6.3481	69	43.8	0.01175
logHs	1	0.5882	68	43.212	0.44311
logD	1	1.5709	67	41.641	0.21008

Table 42: ANOVA and Chi-square for Llobregat with Dates (Threshold= 200 cm)

	Degree of Freedom	Deviance	Residual Degree of Freedom	Residual Deviance	P=(>|Chi|)
NULL	71	50.232			
Season	2	0.4816	69	49.75	0.786
Height	1	0.7133	68	49.037	0.39835
Duration	1	0.8674	67	48.17	0.35167
Season: Height	2	0.725	65	47.445	0.69595
Season: Duration	2	8.5581	63	38.886	0.01386

Table 43: ANOVA and Chi-square for Llobregat with seasons (Threshold= 200 cm)

Season	No Twin	Twin	**Total**
Spring	21	5	26
Fall	55	15	70
Winter	46	18	64
Total	122	38	160

Table 44a: Contingency Table for Llobregat Storms

Season	No Twin	Twin	**Total**
Spring	13	1	14
Fall	31	4	35
Winter	19	3	22
Total	63	8	71

Table 44.b: Contingency Table for Llobregat Storms (Threshold= 200 cm)

Table 44: Contingency Tables for Llobregat

Season	No Twin	Twin	**Total**
Spring	0.13125	0.03125	0.1625
Fall	0.34375	0.09375	0.4375
Winter	0.2875	0.1125	0.4
Total	0.7625	0.2375	1

Table 45.a: Probability Table for Llobregat Storms

Season	No Twin	Twin	**Total**
Spring	0.183099	0.014085	0.197183
Fall	0.43662	0.056338	0.492958
Winter	0.267606	0.042254	0.309859
Total	0.887324	0.112676	1

Table 45.b: Probability Table for Llobregat Storms (Threshold = 200 cm)

Table 45: Probability Tables for Llobregat

Appendix F: Applied "R" Commands

Statistical Inference (Example)

Importing Data

```
LlobregatDates= read.table("LlobregatAllwithDates.txt")

LlobregatSin= LlobregatDates[,2]

LlobregatCos= LlobregatDates[,3]

LlobregatHeight= log10(LlobregatDates[,4])

LlobregatDuration= log10(LlobregatDates[,5])

LlobregatTwin= log10(LlobregatDates[,6])
```

Conditional Density Plots

```
> cdplot= (LlobregatHeight, LlobregatTwin, col= c("red","blue"), main=
"Llobregat Storm Wave Height")

LlobregatTwin= as.factor(LlobregatDates[,6])

cdplot(LlobregatHeight, LlobregatTwin, col= c("red","blue"), main= "Llobregat
Storm Wave Height")

cdplot(LlobregatDuration, LlobregatTwin, col= c("red","blue"), main=
"Llobregat Storm Duration")

coplot(LlobregatHeight ~ LlobregatDuration|LlobregatTwin, main= "Llobregat",
col= c("red", "blue"))
```

Nominal Data Model

```
LlobregatDatesModel= glm(LlobregatTwin ~ LlobregatSin + LlobregatCos +
LlobregatHeight + LlobregatDuration, family = binomial, data= LlobregatDates)

anova(LlobregatDatesModel, test= "Chisq")
```

Ellipse

```
LlobregatEllipse= ellipse(LlobregatDatesModel, which= c(2,3))

plot(LlobregatEllipse, main= "Llobregat Date Ellipse")

abline(h=0,v=0, col= c("red"))
```

Seasons (Categorical Variables) Model

```
LlobregatSeasons= read.table("LlobregatSeason.txt")

LlobregatSeason= as.factor(LlobregatSeasons[,2])

LlobregatSeasonModel= glm(LlobregatTwin~ LlobregatSeason*LlobregatHeight +
LlobregatSeason*LlobregatDuration, family= binomial, data= LlobregatSeasons)

anova(LlobregatSeasonModel, test= "Chisq")
```

Mosaic Plots

```
mosaicplot(LlobregatSeason~ LlobregatTwin, col= c("red","blue"), main=
"Mosaic of Llobregat Twin vs. Season")

plot(LlobregatSeason, LlobregatTwin, col= c("red","blue"))
```

Contingency Tables and Odds Ratio

```
LlobregatCT= table(LlobregatSeason, LlobregatTwin)

oddsratio.fisher(LlobregatCT, y= NULL, conf.level= 0.95, verbose = TRUE)
```